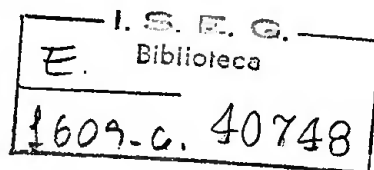


Instituto Superior de Economia e Gestão  
Universidade Técnica de Lisboa

**RESERVADO**



HB822  
N47  
1993

## **MODELOS DE ANÁLISE DE INTERVENÇÃO NO CONTEXTO UNIVARIADO**

**Uma Aplicação à Correção do Indicador do Consumo Corrente**

por

**Renata Maria de Menezes Mesquita**

Lisboa 1993



Instituto Superior de Economia e Gestão  
Universidade Técnica de Lisboa

## **MODELOS DE ANÁLISE DE INTERVENÇÃO NO CONTEXTO UNIVARIADO**

**Uma Aplicação à Correção do Indicador do Consumo Corrente**


por

**Renata Maria de Menezes Mesquita**

Dissertação apresentada como requisito  
parcial para a obtenção do grau de Mestre  
em Matemática Aplicada à Economia e Gestão

1993

# ERRATA

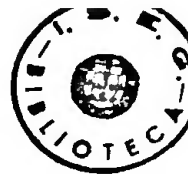
 Pág	Linha	Onde se lê	Deve ler-se
4	nota de pé-de-página [6]	$[\Phi_p(B) \text{ e } \phi_p(B)]$	$[\Phi_p(B^*) \text{ e } \phi_p(B)]$
4	nota de pé-de-página [6]	$[\Theta_q(B) \text{ e } \theta_q(B)]$	$[\Theta_q(B^*) \text{ e } \theta_q(B)]$
11	16 e 18	$\xi_m$	$D_m$
16	21	ordem para	ordem s para
21	nota de pé-de-página [2]	para facilitara	para facilitar a
25	11	estatística teste	estatística do teste
31	18	simplesmente	simplesmente
39	13	$\hat{\beta}A^{-1}\hat{\beta}$	$\hat{\beta}^T A^{-1} \hat{\beta}$
39	14	$\hat{\beta} = (\hat{\beta}_1, \dots, \hat{\beta}_k)$	$\hat{\beta} = (\hat{\beta}_1 \dots \hat{\beta}_k)^T$
39	16	$\hat{\beta}A^{-1}\hat{\beta}$	$\hat{\beta}^T A^{-1} \hat{\beta}$
39	nota de pé-de-página [9]	$\beta=(\beta_1, \dots, \beta_7)$ $\Phi=(\Phi_1, \dots, \Phi_p)$ $\phi=(\phi_1, \dots, \phi_p)$ $\Theta=(\Theta_1, \dots, \Theta_q)$ $\theta=(\theta_1, \dots, \theta_q)$	$\beta=(\beta_1 \dots \beta_7)^T$ $\Phi=(\Phi_1 \dots \Phi_p)^T$ $\phi=(\phi_1 \dots \phi_p)^T$ $\Theta=(\Theta_1 \dots \Theta_q)^T$ $\theta=(\theta_1 \dots \theta_q)^T$
45	21	este efeito	estes efeitos
46	nota de pé-de-página [16]	sujacentes	subjacentes
47	20	Chen e Liu (1993)	Chen e Liu (1993a)
55	16	figuras 11 a 13	figuras 11 a 15
56	7	figuras 14 e 15	figuras 13 e 14
60	16	siginifica	significa
62	3 [equação (4.1)]	$\sum_{H=1}^9$	$\sum_{H=1}^9$
62	3 [equação (4.1)]	$\sum_{H=1}^5$	$\sum_{H=1}^5$

# **ERRATA (cont.)**

Pág	Linha	Onde se lê	Deve ler-se
66	1	presença dois tipos	presença dos dois tipos
66	15	sexta-feiras	sextas-feiras
92	2	heteroescedasticidade dado	heteroescedasticidade é dado
94	2	$\left( \frac{\sum_{n_1} t_i^{(1)} + \frac{1}{\sqrt{k}} \sum_{n_2} t_i^{(2)}}{n_1 + \frac{n_2}{k}} \right)$	$\left( \frac{\sum_{n_1} t_i^{(1)} + \frac{1}{k} \sum_{n_2} t_i^{(2)}}{n_1 + \frac{n_2}{k}} \right)$
94	7	$\frac{e_{n+1}}{\sqrt{k}} - \left( \frac{\sum_{n_1} e_i^{(1)} + \frac{1}{\sqrt{k}} \sum_{n_2} \frac{e_i^{(2)}}{\sqrt{k}}}{n_1 + \frac{n_2}{k}} \right)$	$\frac{e_{n+1}}{\sqrt{k}} - \frac{1}{\sqrt{k}} \left( \frac{\sum_{n_1} e_i^{(1)} + \frac{1}{\sqrt{k}} \sum_{n_2} \frac{e_i^{(2)}}{\sqrt{k}}}{n_1 + \frac{n_2}{k}} \right)$



*Aos meus Pais*



Ao meu orientador da tese, o Professor Doutor Silva Ribeiro, cabe aqui o meu agradecimento pelos comentários e sugestões que efectuou ao longo da elaboração deste trabalho e pela forma como acompanhou a sua realização.

Estou profundamente grata ao meu coordenador de Núcleo no Banco de Portugal, o Dr. Carlos Robalo, pelo seu incentivo e disponibilidade que sempre apresentou, bem como pelas suas sugestões e críticas que se revelaram sempre pertinentes. Estou também grata ao meu colega de gabinete, o Dr. Francisco Dias, e ao Professor Maximiano Pinheiro, com quem as trocas de impressões foram sempre proveitosas e esclarecedoras.

Gostaria, também, de deixar aqui o meu agradecimento ao Dr. Ascensão Costa, do ISEG, pelas várias referências bibliográficas que me facultou e que se revelaram fundamentais para uma melhor compreensão sobre determinados pontos do tema.

O meu agradecimento à JNICT (Junta Nacional de Investigação Científica e Tecnológica) pela atribuição da Bolsa para frequentar o Mestrado (BM/585/90-RO).

Deixo também aqui um obrigado muito especial a vários colegas do Banco e amigos pelo seu apoio, entre os quais não posso deixar de mencionar a Cristina Manteu, Ildeberta Abreu, Ana Cristina Fernandes, Rui Albuquerque, João Cocco, Armando Verdasca, Carmen Sequeira, Cristina Casalinho, Sónia Costa, Isabel Reis e Carla Jardim.

Por último, e em especial, agradeço à minha família, aos meus pais, irmãos e avó, por todo o apoio e força que me deram durante a elaboração desta dissertação.

Quaisquer erros ou omissões são da minha inteira responsabilidade.

## Índice

<b>0 - Introdução.....</b>	<b>1</b>
<b>I - Metodologia da Análise de Intervenção.....</b>	<b>3</b>
I.1 - Modelos de Função Transferência.....	3
I.2 - Modelos de Análise de Intervenção.....	6
I.2.1 - Identificação, Estimação e Avaliação do Diagnóstico dos Modelos de Análise de Intervenção.....	10
I.2.1.1 - Identificação do Modelo de Análise de Intervenção.....	10
I.2.1.2 - Estimação e Avaliação do Diagnóstico do Modelo de Análise de Intervenção.....	15
<b>II - <i>Outliers</i> em Séries Temporais: Procedimento Iterativo de     Detecção e Correção.....</b>	<b>19</b>
II.1 - Os Diversos Tipos de <i>Outliers</i> .....	20
II.2 - Consequências da Presença de <i>Outliers</i> nos Resultados de Modelação e Previsão Box-Jenkins Univariada.....	22
II.3 - Estimação dos Efeitos dos <i>Outliers</i> Assumindo que os Parâmetros do Modelo ARIMA são Conhecidos.....	23
II.4 - Procedimento de Detecção e Estimação dos Efeitos de Múltiplos <i>Outliers</i> quando os Parâmetros do Modelo ARIMA são Desconhecidos.....	26

<b>III - Efeito "Dias da Semana" e Efeito "Páscoa": Procedimentos de Detecção e Correção.....</b>	<b>30</b>
III.1 - Efeito "Dias da Semana" .....	31
III.1.1 - Correção do Efeito "Dias da Semana" .....	31
III.1.2 - Detecção do Efeito "Dias da Semana" .....	38
III.2 - Efeito "Páscoa": Detecção e Correção.....	41
 <b>IV - Aplicação da Metodologia da Análise de Intervenção à Correção do Indicador do Consumo Corrente.....</b>	 <b>47</b>
IV.1 - O Indicador do Consumo Corrente.....	48
IV.2 - Primeira Hipótese de Correção do Indicador do Consumo Corrente.....	50
IV.2.1 - Definição das Variáveis de Intervenção.....	50
IV.2.1.1 - Efeito "Dias da Semana" .....	51
IV.2.1.2 - Efeito "Páscoa" .....	52
IV.2.1.3 - Efeito "Rendimento" .....	55
IV.2.1.4 - Efeito "Novo Hipermercado" .....	57
IV.2.2 - Resultados Empíricos.....	61



IV.3 - Segunda Hipótese de Correção do Indicador do Consumo Corrente.....	71
IV.3.1 - Construção do "Novo" Indicador do Consumo Corrente.....	71
IV.3.2 - Correção dos Efeitos "Dias da Semana", "Páscoa" e "Rendimento".....	79
V-Considerações Finais.....	90
Apêndice: Definição do Critério de Entrada das Grandes Superfícies no Cálculo do Índice do Consumo Corrente.....	91
Anexo de Figuras e Tabelas.....	97
- Figuras e Tabelas da Primeira Hipótese de Correção.....	97
- Figuras e Tabelas da Segunda Hipótese de Correção.....	113
Bibliografia.....	128

## Índice de Figuras e Tabelas

### ♦ Primeira Hipótese de Correção do Indicador do Consumo Corrente

#### FIGURAS

1 - Índice do Consumo Corrente (ICC).....	98
2 - Logaritmo do Índice do Consumo Corrente (LICC).....	98
3 - Taxas de Variação Homólogas do Índice do Consumo Corrente.....	98
4 - Efeito "Dias da Semana": Vendas Nominais Diárias na Semana de 8-14 de Abril de 1991.....	99
5 - Efeito "Dias da Semana": Vendas Nominais Diárias na Semana de 17-23 de Junho de 1991.....	99
6 - Efeito "Dias da Semana": Vendas Nominais Diárias na Semana de 16-22 de Setembro de 1991.....	99
7 - Efeito "Dias da Semana": Vendas Nominais Diárias na Semana de 11-17 de Novembro de 1991 .....	99
8 - Efeito "Páscoa": Vendas Nominais Diárias de Abril de 1992.....	100
9 - Efeito "Páscoa": Coeficientes de Ponderação Diários do Efeito "Expansivo" nas Vendas.....	53
10 - Efeito "Páscoa": Coeficientes de Ponderação Diários do Efeito "Depressivo" nas Vendas.....	54
11 - Efeito "Rendimento": Vendas Nominais Diárias de Janeiro de 1991.....	101
12 - Efeito "Rendimento": Vendas Nominais Diárias de Fevereiro de 1991.....	101
13 - Efeito "Rendimento": Vendas Nominais Diárias de Maio de 1991.....	102
14 - Efeito "Rendimento": Vendas Nominais Diárias de Junho de 1991.....	102
15 - Efeito "Rendimento": Vendas Nominais Diárias de Agosto de 1991.....	102
16 - Função de Autocorrelação da Série Residual do Modelo (4.1).....	105
17 - Função de Autocorrelação Parcial da Série Residual do Modelo (4.1).....	105
18 - Série Residual do Modelo (4.1) Ajustada de <i>Outliers</i> .....	105
19 - Série Residual do Modelo "Final" Ajustada de <i>Outliers</i> .....	106
20 - Função de Autocorrelação da Série Residual do Modelo "Final" .....	107
21 - Função de Autocorrelação Parcial da Série Residual do Modelo "Final" .....	107
22 - Função de Autocorrelação da Série LICC.....	108
23 - Função de Autocorrelação da Série $(1-B^{1/2})LICC$ .....	108
24 - Função de Autocorrelação da Série $(1-B)(1-B^{1/2})LICC$ .....	108
25 - Taxas de Variação Homólogas do Índice do Consumo Corrente Publicadas e Corrigidas.....	109

## TABELAS

1.1 - Efeito "Páscoa": Coeficientes de Ponderação Diários do Efeito "Expansivo" nas Vendas.....	54
1.2 - Efeito "Páscoa": Coeficientes de Ponderação Diários do Efeito "Depressivo" nas Vendas.....	54
- Determinação do Efeito "Rendimento" .....	103
1.3.A - Especificação "Final" do Modelo (4.1).....	64
1.3.B - Detecção e Ajustamento de <i>Outliers</i> .....	65
1.3.C - Coeficientes de Autocorrelação da Série Residual.....	65
1.3.D - Coeficientes de Correlação das Estimativas dos Parâmetros.....	106
1.4 - Impacte da Abertura de um Novo Hipermercado sobre o Indicador do Consumo Corrente.....	70
1.5 - Índice Publicado e Índice Corrigido e Taxas de Variação Homólogas Publicadas e Corrigidas.....	110

### ♦ Segunda Hipótese de Correção do Indicador do Consumo Corrente

## FIGURAS

26 - Índice do Consumo Corrente (ICC2).....	120
27 - Logaritmo do Índice do Consumo Corrente (LICC2).....	120
28 - Taxas de Variação Homólogas do Índice Publicado (ICC) e do "Novo" Índice do Consumo Corrente (ICC2).....	120
29 - Função de Autocorrelação da Série LICC2.....	121
30 - Função de Autocorrelação da Série $(1-B^{12})LICC2$ .....	121
31 - Função de Autocorrelação da Série Residual do Modelo (4.12).....	122
32 - Função de Autocorrelação Parcial da Série Residual do Modelo (4.12).....	122
33 - Série Residual do Modelo (4.12) Ajustada de <i>Outliers</i> .....	122
34 - Função de Autocorrelação da Série Residual do Modelo "Final" .....	123
35 - Função de Autocorrelação Parcial da Série Residual do Modelo "Final" .....	123
36 - Série Residual do Modelo "Final" Ajustada de <i>Outliers</i> .....	124
37 - Taxas de Variação Homólogas do ICC2 Não Corrigidas e Corrigidas.....	125
38 - Taxas de Variação Homólogas Corrigidas: Hipótese 2 <i>versus</i> Hipótese 1.....	126
39 - Taxas de Variação Homólogas do ICC2 — Correção pelo X11ARIMA.....	127
40 - Taxas de Variação Homólogas do ICC2 — Correção pelo X11ARIMA e pelo Modelo de Intervenção.....	127

## TABELAS

2.1 - Taxas de Variação Homólogas: Publicadas (do ICC), do "Novo" Índice (do ICC2) e Corrigidas.....	114
2.2 - "Novo" Índice (ICC2) e "Novo" Índice Corrigido.....	117
2.3.A - Especificação "Final" do Modelo (4.12).....	82
2.3.B - Detecção e Ajustamento de Outliers.....	82
2.3.C - Coeficientes de Autocorrelação da Série Residual.....	83
2.3.D - Coeficientes de Correlação das Estimativas dos Parâmetros da Especificação "Final" do Modelo (4.12).....	124
2.4 - Especificação "Final" do Modelo Restrito (4.14).....	84
2.5.A - Especificação "Final" do Modelo (4.15).....	86
2.5.B - Coeficientes de Correlação das Estimativas dos Parâmetros da Especificação "Final" do Modelo (4.15).....	87
2.6 - Variáveis $PASP_t$ e $PASN_t$ : Proposta de Escrivá e Sastre para o Efeito "Páscoa" .....	88

## 0 - INTRODUÇÃO

Em aplicações empíricas, realizadas no âmbito das sucessões cronológicas, é frequente detectarem-se comportamentos atípicos no andamento das séries temporais em estudo. Estes comportamentos tendem frequentemente a estar associados à ocorrência de fenómenos de natureza exógena e determinística, cujos efeitos sobre a série temporal devem ser quantificados por forma a serem tidos em conta no respectivo processo de modelação. Estudos efectuados com o objectivo de avaliar as consequências de se ignorarem os efeitos desses fenómenos sobre os resultados do processo de modelação e previsão Box-Jenkins demonstraram que estes poderão ficar seriamente comprometidos, atendendo aos potenciais enviesamentos que esses efeitos produzem, quer sobre as estatísticas com base nas quais se procede à identificação do modelo quer sobre os próprios estimadores dos parâmetros do modelo ARIMA.

Relativamente ao momento em que ocorrem, existem situações em que se conhece à partida a data de ocorrência do "choque" exógeno, enquanto que, em outras situações, esse dado é desconhecido. A metodologia de tratamento de séries temporais afectadas por "choques" pertencentes à primeira categoria, ou seja, por "choques" cujo momento de ocorrência é conhecido, foi proposta por Box e Tiao (1975) tendo ficado conhecida na literatura por análise de intervenção. Estes autores estendem os modelos univariados Box-Jenkins à inclusão de variáveis determinísticas, habitualmente designadas por variáveis de intervenção, cujo objectivo é precisamente o de captarem os efeitos desses "choques" sobre a série temporal. No que respeita aos "choques" cujo momento de ocorrência é desconhecido, o seu estudo insere-se na chamada análise de *outliers*. Atendendo a que estes fenómenos podem ser vistos como intervenções que têm como particularidade o facto de não se conhecer à partida o momento em que ocorrem, os procedimentos desenvolvidos, no âmbito da análise de *outliers*, para detectar e corrigir os *outliers* em séries temporais podem considerar-se como alternativos mas, ao mesmo tempo, complementares dos procedimentos subjacentes à análise de intervenção, pelo que as duas metodologias se encontram intimamente ligadas.

Pretende-se, neste trabalho, apresentar e, através de uma aplicação, exemplificar o processo de modelação de séries temporais "contaminadas" por fenómenos de carácter exógeno e determinístico cujo momento de ocorrência encontra-se bem localizado no tempo.

Deste modo, a partir da relação que é possível estabelecer-se entre a metodologia da função transferência e da análise de intervenção, começa-se, no primeiro capítulo, por abordar, do ponto de vista teórico, os modelos de intervenção, explicitando cada um dos passos subjacentes ao respectivo processo de modelação, mais concretamente, as etapas de identificação, estimação e avaliação do diagnóstico.

No segundo capítulo, e tendo presente a relevância da análise de *outliers* como metodologia complementar da análise de intervenção, procede-se a uma descrição sumária de um dos procedimentos iterativos de detecção e ajustamento de observações atípicas em séries temporais. Este procedimento, desenvolvido por Chen e Liu em finais da década de 80, encontra-se incorporado num dos módulos do programa SCA (*Scientific Computing Associates*), *software* que serviu de base à aplicação empírica realizada. Neste capítulo são também referidos os diversos tipos de *outliers*, usualmente considerados na literatura, bem como as consequências que se produzem ao nível dos resultados do processo de modelação e previsão Box-Jenkins, decorrentes do facto de se ignorar a presença destes fenómenos nas séries temporais. Em seguida, no capítulo III, abordam-se os diferentes procedimentos de detecção e correcção dos efeitos "dias da semana" e "Páscoa", designados conjuntamente por efeitos de "calendário". Embora o conceito de intervenção surja geralmente associado ao de "choque" exógeno que ocorre pontualmente no tempo, considerou-se adequado enquadrar estes dois tipos de fenómenos, de natureza sistemática, no contexto dos modelos de intervenção, dado o carácter determinístico dos efeitos que produzem sobre as séries temporais.

No capítulo IV, e com base nos conceitos e procedimentos apresentados nos capítulos anteriores, aplica-se a metodologia da análise de intervenção ao indicador do consumo corrente, de modo a corrigi-lo dos efeitos de um conjunto de fenómenos que tendem a pôr em causa sua utilidade para efeitos de análise de conjuntura. Estes fenómenos foram designados por efeito "dias da semana", efeito "Páscoa", efeito "rendimento" e efeito "novo hipermercado". Nesta aplicação foram consideradas duas hipóteses de correcção. Na primeira, o indicador do consumo corrente é corrigido dos quatro efeitos acima mencionados, enquanto que, na segunda, procura-se eliminar previamente o efeito "novo hipermercado", através da reconstrução do próprio indicador, após a qual se aplica a metodologia da análise de intervenção no intuito de corrigir o "novo" indicador dos restantes três efeitos. Por último, são tecidas algumas considerações finais sobre a aplicabilidade desta metodologia, tendo por base os resultados empíricos obtidos.

# I - METODOLOGIA DA ANÁLISE DE INTERVENÇÃO

## I.1 - Modelos de Função Transferência

A metodologia da análise de intervenção surge enquadrada nos desenvolvimentos que se verificaram, no contexto da análise de séries temporais<sup>[1]</sup>, após ter ficado estabelecida a relevância dos modelos ARIMA<sup>[2]</sup> univariados desenvolvidos por Box e Jenkins em finais da década de 60, início da década de 70. A necessidade de um maior aperfeiçoamento e generalização dos modelos ARIMA univariados, dadas as potencialidades que apresentavam como alternativa à modelação econométrica para efeitos de previsão no curto prazo, propulsionou a sua extensão ao âmbito multivariado. Em vez de na especificação do modelo se considerarem apenas os valores passados da variável em estudo, seria também incluído um conjunto adicional de variáveis que, por estarem correlacionadas com o andamento desse mesmo fenómeno, permitiam melhorar a capacidade preditiva do respectivo modelo univariado.

É neste contexto que se inserem os modelos de função transferência (daqui em diante designados por MFT), também desenvolvidos por Box e Jenkins, cuja terminologia e muitos dos conceitos teóricos que lhes estão subjacentes se podem aplicar apropriadamente aos modelos de análise de intervenção, dada a relação estreita entre as duas metodologias. Como mais adiante será referido, a análise de intervenção pode ser considerada como um caso particular dos MFT, pelo que partindo da metodologia subjacente a estes últimos facilmente se derivam os resultados teóricos da análise de intervenção<sup>[3]</sup>.

---

<sup>[1]</sup> Ao longo deste trabalho utilizar-se-á indistintamente as designações de "série" temporal e de "sucessão" cronológica para, em rigor, querer designar "sucessão" temporal ou cronológica. Esta deve ser entendida como uma realização de um processo estocástico. Todavia, uma vez que a designação de "série" temporal se encontra vulgarizada, não estando, portanto, em causa a possibilidade de confusão com o conceito matemático de "série", optou-se por utilizar indistintamente as duas designações.

<sup>[2]</sup> Em terminologia inglesa, *autoregressive integrated moving average processes*.

<sup>[3]</sup> Neste trabalho far-se-á uma apresentação resumida dos MFT, realçando apenas os conceitos fundamentais, relevantes para a abordagem da análise de intervenção. Um maior desenvolvimento da metodologia dos MFT pode ser encontrada em Box e Jenkins (1976), Müller (1990), Wei (1990) e Murteira, Müller e Turkman (1993).

Seja, então,  $\{y_t: t=0,1,2,\dots\}$  uma série temporal cujo comportamento se pretende modelar e suponha-se que o processo gerador de  $\{y_t\}$  se compõe de duas componentes distintas não correlacionadas entre si: uma componente sistemática, que representamos por  $f(\mathbf{k}, \xi_t, t)$ , e uma componente não sistemática ou, mais precisamente, uma componente ruído, que representamos por  $N_t$ , cujo comportamento pode ser modelado através de um processo ARMA<sup>[4]</sup>. Combinando de uma forma aditiva estas duas componentes tem-se o modelo:

$$y_t = f(\mathbf{k}, \xi_t, t) + N_t$$

onde  $y$ , habitualmente denominada variável output, corresponde à série original ( $Y$ ) ou a alguma sua transformação (por exemplo,  $y = \ln(Y)$ ), e  $f(\mathbf{k}, \xi_t, t)$  é uma função composta por um conjunto de variáveis exógenas independentes,  $\xi_t$ , também designadas variáveis inputs, por uma componente determinística na variável tempo,  $t$ , e por um vector de parâmetros desconhecidos,  $\mathbf{k}$ .

Admite-se habitualmente que o processo residual ou o noise da série  $y_t$ , traduzido por  $\{N_t = y_t - f(\mathbf{k}, \xi_t, t): t=0,1,2,\dots\}$ , verifica as propriedades  $E(N_t)=0 \forall t$  e  $\text{Cov}(N_t, \xi_{tm})=0 \forall t, m$  e segue um processo ARMA(p,q)(P,Q)<sub>s</sub><sup>[5]</sup> estacionário e invertível<sup>[6]</sup> do tipo:

$$\Phi_P(B^s) \phi_p(B) N_t = \Theta_Q(B^s) \theta_q(B) \varepsilon_t$$

onde  $\Phi_P(B^s)=(1 - \Phi_1 B^s - \Phi_2 B^{2s} - \dots - \Phi_P B^{Ps})$  e  $\phi_p(B)=(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)$  são, respectivamente, os polinómios auto-regressivos das componentes sazonal e não sazonal,  $\Theta_Q(B^s)=(1 - \Theta_1 B^s - \Theta_2 B^{2s} - \dots - \Theta_Q B^{Qs})$  e  $\theta_q(B)=(1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q)$  são os polinómios médias móveis das componentes sazonal e não sazonal, respectivamente, e  $\{\varepsilon_t\}$  é uma

---

<sup>[4]</sup> Autoregressive moving average process, em terminologia inglesa.

<sup>[5]</sup> Esta expressão define um modelo ARMA multiplicativo para uma sucessão cronológica com sazonalidade (sendo  $s$  a periodicidade da série).  $P$  e  $p$  são, respectivamente, as ordens dos factores auto-regressivos da componente sazonal e não sazonal.  $Q$  e  $q$  são as ordens dos factores médias móveis das componentes sazonal e não sazonal, respectivamente.

<sup>[6]</sup> As condições de estacionaridade e de invertibilidade do processo ARMA pressupõem, respectivamente, que os polinómios auto-regressivos  $[\Phi_P(B)$  e  $\phi_p(B)]$  e os polinómios médias móveis  $[\Theta_Q(B)$  e  $\theta_q(B)]$  tenham as raízes fora do círculo unitário.



sucessão de variáveis normais i.i.d.<sup>[7]</sup> de média zero e variância  $\sigma^2$ , ou seja, um processo gaussiano do tipo ruído branco.

Os efeitos das variáveis exógenas  $\{\xi_m: m=1,...,k\}$  sobre a variável *output* podem ser modelados admitindo uma estrutura dinâmica do seguinte tipo:

$$f(k, \xi_t, t)^{[8]} = f(\delta, \omega, \xi) = C + \sum_{m=1}^k v_m(B) \xi_{tm} = C + \sum_{m=1}^k \frac{\omega_m(B) B^{b_m}}{\delta_m(B)} \xi_{tm} \quad (1.1)$$

onde C representa uma constante e  $v_m(B)$  a especificação linear da função transferência para a variável exógena  $\xi_m$  que pretende captar a relação dinâmica entre esta variável *input* e a variável *output* y, em presença de outras variáveis explicativas. É, geralmente, descrita por um polinómio no operador de desfasamento  $B^{[9]}$  ou, mais concretamente, por uma função de resposta dinâmica<sup>[10]</sup> da forma:  $v_m(B) = v_{0m} + v_{1m}B + v_{2m}B^2 + v_{3m}B^3 + \dots$ , em que  $\{v_m, m=0,1,2,\dots\}$  é o conjunto dos pesos da função transferência.

A expressão  $\omega_m(B)B^{b_m}/\delta_m(B)$  constitui uma aproximação racional, ou melhor, uma representação parcimoniosa da forma linear da função transferência  $v_m(B)$ , tal que:

$$\omega_m(B) = (\omega_{0m}B^0 + \omega_{1m}B^1 + \dots + \omega_{s_m m}B^{s_m})$$

$$\delta_m(B) = (1 + \delta_{1m}B + \dots + \delta_{r_m m}B^{r_m})$$

são polinómios no operador de desfasamento B de ordem  $s_m$  e  $r_m$ , respectivamente. Os coeficientes do primeiro polinómio medem os efeitos iniciais da variável exógena  $\xi_m$  sobre a variável *output* y, enquanto que os do segundo traduzem os efeitos permanentes. O

<sup>[7]</sup> Independentes e identicamente distribuídas.

<sup>[8]</sup> Para simplificar a notação ignora-se daqui em diante a variável t, assumindo que a tendência determinística ou é nula ou, caso exista, encontra-se incorporada nas variáveis exógenas.

<sup>[9]</sup> O operador de desfasamento B é definido pela transformação:  $B^a X_t = X_{t-a}$ .

<sup>[10]</sup> Estes modelos são muitas vezes denominados modelos de resposta dinâmica, atendendo a que uma variação em uma das variáveis exógenas tende a gerar um conjunto de efeitos sobre a variável *output* que se distribuem ao longo do tempo.

expoente  $b_m$  mede o desfasamento com que  $\xi_m$  influencia a variável  $y$ , ou seja, o atraso na resposta<sup>[11]</sup>. Assume-se que todas as raízes dos polinómios  $\omega_m(B)$  e  $\delta_m(B)$  estão fora do círculo unitário, condição que garante a estabilidade da função transferência, no sentido de que uma variação finita na variável *input* induz uma variação também finita no *output*.

O modelo de função transferência é obtido da combinação da componente sistemática, dada pela expressão (1.1), com a componente residual, ou seja,

$$y_t = C + \sum_{m=1}^k v_m(B) \xi_{tm} + N_t = C + \sum_{m=1}^k \frac{\omega_m(B) B^{b_m}}{\delta_m(B)} \xi_{tm} + N_t \quad (1.2)$$

## I.2 - Modelos de Análise de Intervenção

Em termos simplificados, pode-se afirmar que a análise de intervenção constitui basicamente um caso particular do modelo de função transferência, em que as variáveis *input*, em vez de aleatórias ou de serem realizações de processos estocásticos, assumem um carácter determinístico, sendo denominadas variáveis de intervenção.

A introdução deste tipo de variáveis nos modelos de séries cronológicas visa captar o impacto, sobre o comportamento de uma determinada variável, de um conjunto de fenómenos cujos momentos de ocorrência no tempo são conhecidos. Estes fenómenos de carácter exógeno e perfeitamente identificáveis no tempo estão, na sua maioria, associados a acontecimentos diversos, como sejam situações de conflito (sendo as guerras, greves e revoluções exemplos típicos), alterações de legislação (por exemplo, a introdução ou agravamento de um determinado imposto) ou de regimes de política monetária/cambial, variações na composição do calendário (efeito "dias da semana" e efeito "festas móveis"), etc. Como se pode inferir, são inumeráveis os casos em que será de todo pertinente incorporar os efeitos destes fenómenos exógenos na modelação univariada da série temporal em estudo. A não incorporação destes "choques", no modelo ARIMA univariado,

---

<sup>[11]</sup> Na literatura é habitual designá-lo por "dead time".

poderá ter graves repercussões quer no que se refere aos estimadores dos seus parâmetros quer no que respeita à previsão efectuada com base nesse modelo<sup>[12]</sup>. É de referir que, embora a aplicação empírica da análise de intervenção tenha tido origem na necessidade de quantificar os efeitos da ocorrência de fenómenos do tipo acima indicados<sup>[13]</sup> e, actualmente, continue a ser esse o principal objectivo que lhe está subjacente, a sua aplicabilidade encontra-se perfeitamente justificada em muitas outras situações, como por exemplo, de alteração na definição da própria variável em estudo<sup>[14]</sup>.

As variáveis de intervenção são, na maior parte dos casos, variáveis binárias<sup>[15]</sup>, habitualmente conhecidas por variáveis *dummy* ( $D_j$ )<sup>[16]</sup>. Em particular, é usual admitirem-se dois tipos de variáveis de intervenção<sup>[17]</sup>. A variável "impulso" (*pulse variable*), que corresponde a uma intervenção que ocorre no momento  $t=T$  e cuja actuação cessa nesse preciso momento, podendo a sua estrutura ser definida da seguinte forma:

$$D_t = P_t(T) = \begin{cases} 1, & \text{para } t = T \\ 0, & \text{para } t \neq T \end{cases}$$

---

<sup>[12]</sup> No capítulo II, referente à detecção e ajustamento de *outliers*, abordar-se-á esta questão.

<sup>[13]</sup> Tem-se como exemplo o artigo de Box e Tiao (1975), considerado como um dos artigos pioneiros na aplicação da análise de intervenção, em que se procura quantificar o impacto ao nível ambiental e económico de um conjunto de intervenções.

<sup>[14]</sup> Algumas séries económicas, usualmente utilizadas em trabalhos de econometria aplicada, estão sujeitas a este tipo de alteração, ou seja, a situações de quebra de série. Como exemplos, refiram-se as séries do emprego, cujas modificações estão associadas à existência de distintos inquéritos ao emprego, e as séries dos agregados monetários que sofreram alterações na sua composição após o surgimento de novos instrumentos financeiros.

<sup>[15]</sup> Exceptuam-se, como mais à frente se verá, as variáveis de intervenção respeitantes às variações na composição do calendário.

<sup>[16]</sup> Existem autores — e.g. Espasa (1982) — que consideram a metodologia da análise de intervenção como uma generalização da "análise com variáveis artificiais" dado que a primeira flexibiliza esta última ao permitir que cada variável artificial (*dummy*) seja afectada por um filtro ARMA do tipo  $\omega_m(B)\delta_m(B)$ .

<sup>[17]</sup> Utiliza-se aqui a mesma denominação atribuída por Muteira, Müller e Turkman (1993) para designar estes dois tipos de variáveis de intervenção.

e a variável "degrau" (step variable), que está associada a uma intervenção que ocorre no momento  $t=T$  e que permanece actuante, a um nível constante, a partir desse momento do tempo. Este tipo de variável admite uma estrutura da forma:

$$D_t = S_t(T) = \begin{cases} 0, & \text{para } t < T \\ 1, & \text{para } t \geq T \end{cases}$$

Entre estes dois tipos de variáveis de intervenção é possível estabelecer a seguinte relação:

$$P_t(T) = \Delta S_t(T), \text{ em que } \Delta \text{ é o operador diferença, isto é, } \Delta = (1 - B)$$

Um outro tipo de variável, por vezes referido na literatura (e.g, Mills (1990)) e que se tem revelado adequado para representar determinadas formas de intervenções, é a extended pulse variable. Este tipo de variável admite uma estrutura que, de certo modo, se pode considerar intermédia das estruturas das variáveis "impulso" e "degrau", ou seja

$$D_t = \text{EXTP}_t(T_1, T_2) = \begin{cases} 0, & \text{para } t < T_1 \\ 1, & \text{para } T_1 \leq t \leq T_2 \\ 0, & \text{para } t > T_2 \end{cases}$$

onde  $T_1$  e  $T_2$  representam, respectivamente, o primeiro e último momentos de ocorrência da intervenção. Repare-se que a *extended pulse variable* pode ser interpretada como uma sequência de  $(T_2 - T_1 + 1)$  "choques" exógenos do tipo "impulso"<sup>[18]</sup> ou como uma intervenção do tipo "degrau" truncada no momento  $t=T_2$ .

---

[18]  $\text{EXTP}_t(T_1, T_2) = \sum_{i=0}^{T_2-T_1} P_t(T_1+i) = (1+B+\dots+B^{T_2-T_1})P_t(T_1)$  .

Espasa (1982) faz ainda referência a um outro tipo de variável de intervenção que visa representar situações de alteração na tendência da série temporal após a ocorrência de um determinado "choque" exógeno e à qual denomina variável tendência [ $T_t(T)$ ]. Como o seu nome sugere, esta variável assume o valor zero até ao momento de ocorrência da intervenção ( $t=T$ ), tomando a partir dessa data valores 1,2,3,... . Dado que também é possível estabelecer uma relação simples entre as variáveis  $P_t(T)$ ,  $S_t(T)$  e  $T_t(T)$ <sup>[19]</sup>, Espasa (1982) acaba por atribuir uma maior relevância às variáveis "impulso" e "degrau".

Tendo presente o MFT multivariado definido na regressão (1.2) e a relação que existe entre este tipo de modelos e os modelos de análise de intervenção, pode-se propor a seguinte especificação para o **modelo de análise de intervenção**:

$$y_t = C + \sum_{m=1}^k v_m(B) D_{tm} + N_t = C + \sum_{m=1}^k \frac{\omega_m(B) B^{b_m}}{\delta_m(B)} D_{tm} + N_t \quad (1.3)$$

em que se supõe a ocorrência de  $k$  intervenções que podem assumir qualquer um dos quatro tipos de variável atrás indicados consoante a natureza e duração da intervenção.

Relativamente aos modelos de análise de intervenção, e antes de prosseguir com a identificação e estimação destes modelos, existem dois aspectos particulares que merecem referência dado que se distinguem dos MFT. O primeiro aspecto está relacionado com o comportamento estocástico da componente residual  $N_t$  que, no âmbito dos modelos de intervenção, poderá seguir um processo ARIMA( $p,d,q$ )( $P,D,Q$ )<sup>[20]</sup> do tipo  $\Phi_P(B^s)\phi_p(B)\Delta_s^D\Delta^dN_t = \Theta_Q(B^s)\theta_q(B)\varepsilon_t$ <sup>[21]</sup>, uma vez que nestes modelos as variáveis exógenas são determinísticas. O segundo aspecto diz respeito ao polinómio  $\delta_m(B)$  que, no contexto dos modelos de análise de intervenção, pode admitir raízes sobre o círculo unitário, ao

---

[19] Tem-se  $T_t(T) = \frac{1}{(1-B)}S_t(T) = \frac{1}{(1-B)(1-B)}P_t(T)$  .

[20] Onde  $P$ ,  $p$ ,  $Q$ ,  $q$  e  $s$  têm o mesmo significado que o referido na nota de pé-de-página [5] e  $D$  e  $d$  são o número de diferenciações de ordem  $s$  e de ordem um necessário para tornar a série estacionária. Tomou-se aqui o conceito fraco de estacionaridade.

[21] Onde  $\{\varepsilon_t\}$  continua a ser uma série ruído branco com uma distribuição  $N(0,\sigma_\varepsilon^2)$ .

contrário do que sucede nos MFT. A existência de raízes unitárias em  $\delta_m(B)$  pode ser interpretada como uma resposta que cresce linearmente no tempo, ou melhor, como um efeito que não se esgota ao fim de um número finito de períodos de tempo (veja-se nota-de-pé de página [19]).

### I.2.1 - Identificação, Estimação e Avaliação do Diagnóstico dos Modelos de Análise de Intervenção

Na abordagem da análise de intervenção, tal como no processo de modelação Box-Jenkins no contexto univariado e nos MFT, as etapas de identificação, estimação e avaliação do diagnóstico realizam-se segundo uma lógica sequencial e iterativa que termina quando se obtiver uma ou várias especificações que verifiquem os critérios definidos na etapa do controlo do diagnóstico.

#### I.2.1.1 - Identificação do Modelo de Análise de Intervenção

Nesta etapa é mais correcto admitirem-se dois "tipos" de identificação: uma "pseudo" identificação que respeita à identificação da função transferência  $v_m(B)$  associada a cada variável de intervenção, ou seja, à determinação das ordens dos polinómios  $\omega_m(B)$  e  $\delta_m(B)$  ( $s_m$  e  $r_m$ , respectivamente), e uma identificação propriamente dita<sup>[22]</sup> que diz respeito à modelação ARIMA da série  $N_t$ .

Ao contrário do que acontece nos MFT em que, nesta etapa, a estatística de correlação cruzada desempenha um papel fulcral, nos modelos de intervenção a **identificação da função transferência** realiza-se recorrendo a um dos três procedimentos a seguir indicados, cujos pressupostos básicos são o conhecimento quer do momento de ocorrência quer da natureza de cada uma das intervenções (se é uma função "degrau", "impulso", "tendência" ou uma função do tipo *extended pulse*):

---

<sup>[22]</sup> Entenda-se como uma identificação que utiliza os mesmos procedimentos que estão subjacentes à primeira etapa da modelação univariada Box-Jenkins.

a) a partir do cronograma da série original procurar visualizar os efeitos das intervenções na série temporal. Este procedimento tem o inconveniente de se tornar impraticável perante a ocorrência de vários choques exógenos cujos efeitos, ao se propagarem no tempo, entrecruzam-se, tornando extremamente difícil, se não impossível, identificá-los correctamente numa base individual.

b) especificar a forma funcional dos polinómios  $\omega_m(B)$  e  $\delta_m(B)$  com base numa hipótese de partida julgada adequada para traduzir as repercussões da intervenção  $D_m$  ( $m=1,...,k$ ) sobre a variável *output*  $y^{[23]}$ . Assim, por exemplo, se ocorrer uma intervenção do tipo "impulso" no momento  $t=T$ , cujo impacte sobre  $y_t$  se espera que se esgote nesse momento, ter-se-á  $\omega_m(B)=\omega_0$  e  $\delta_m(B)=1$ . Se, por outro lado, se admitir que o efeito inicial ocorre só ao fim de um período e se propaga no tempo com um comportamento exponencialmente amortecido, então será aconselhável tomar  $\omega_m(B) = \omega_0$ ,  $b_m = 1$  e  $\delta_m(B)=1-\delta_m B$  (com  $0 < \delta_m < 1$ ).

c) Determinados autores, entre eles Liu e Hanssens (1982), propõem a aplicação do método LTF (*linear transfer function*), cujo primeiro passo consiste na especificação de uma função transferência, para cada variável de intervenção  $\xi_m$  ( $m=1,...,k$ ), truncada no desfasamento  $l_m$  (sendo  $l_m$  um número elevado). Neste caso, começa-se por se estimar o modelo (1.3) tal que à variável  $\xi_m$  estará associada a seguinte função transferência:  $v_m(B)=v_{0m} + v_{1m}B + v_{2m}B^2 + v_{3m}B^3+...+ v_{l_m}B^{l_m}$ . Desta forma, ao proceder-se a uma primeira avaliação da qualidade estatística do modelo, eliminar-se-iam os desfasamentos cujos parâmetros associados vieram estatisticamente não significativos. Atendendo à dimensão das sucessões cronológicas com que habitualmente se trabalha em estudos empíricos, esta abordagem do "geral para o particular" tem como limitação o reduzido número de graus de liberdade com que se fica quando tanto  $l_m$  ( $m=1,...,k$ ) como  $k$  são elevados.

---

[23] Box e Tiao (1975), Espasa (1982), Müller (1990) e Murteira, Müller e Turkman (1993) apresentam as especificações das funções transferência habitualmente aplicadas em estudos empíricos e as respectivas representações gráficas.

A identificação do modelo ARIMA para a componente ruído ( $N_t$ ) traduz-se na especificação de um modelo para a série original expurgada dos efeitos das intervenções, ou seja, para a série não "contaminada" por intervenções. No contexto dos modelos de análise de intervenção esta etapa realiza-se recorrendo aos procedimentos e estatísticas habituais subjacentes à fase de identificação dos modelos Box-Jenkins<sup>[24]</sup>, sendo geralmente consideradas três alternativas para a modelação do referido modelo:

i) especificação do modelo ARIMA utilizando apenas os dados do período que antecede a ocorrência da primeira intervenção (em  $t=T_1$ ), ou seja,  $\{y_t: t < T_1\}$  ou, então, utilizando os dados posteriores ao momento em que se esgotaram os efeitos da última intervenção, dependendo do número de observações disponíveis em cada um dos sub-períodos.

ii) especificação do modelo ARIMA utilizando a totalidade dos dados da série temporal, com excepção dos que respeitam aos períodos em que ocorrem as intervenções e se esgotam os seus efeitos. Por exemplo, numa situação em que apenas um período da série temporal está "contaminado" por intervenções, o modelo ARIMA para  $N_t$  será identificado, segundo esta alternativa, com base numa "nova" série temporal que resulta da "colagem" dos dois sub-períodos referidos<sup>[25]</sup>. Um contra-argumento à aplicação desta hipótese é o de que a eliminação de observações intermédias poderá pôr em causa a correcta identificação do modelo, uma vez que produzirá um enviesamento das estatísticas a partir das quais se procede à sua identificação, nomeadamente, das funções de autocorrelação (FAC) e autocorrelação parcial (FACP). Na realidade, os coeficientes de autocorrelação da "nova" série tenderão a ser subestimados, relativamente aos da série original, sendo que essa subestimação será tanto mais elevada quanto maior o número de observações eliminadas<sup>[26]</sup>. Além disso, as consequências de se eliminarem observações

---

<sup>[24]</sup> Mais concretamente, a partir das funções de autocorrelação - FAC - e autocorrelação parcial - FACP. Em estudos mais recentes, é usual calcular-se também a *extended autocorrelation function* - EACF - que, segundo diversos autores, permite uma identificação mais rápida das ordens dos polinómios auto-regressivos e de médias móveis.

<sup>[25]</sup> Como exemplo, suponha-se que  $T_1$  é o momento de ocorrência da primeira intervenção e que  $T'$  é o momento do tempo que delimita superiormente o período em que já se deram todas as intervenções e já se esgotaram os seus efeitos. A "nova" série temporal com base na qual se procederá à identificação do modelo ARIMA para  $N_t$  será  $y_t' = \{y_t: t < T_1\} \cup \{y_t: t \geq T'\}$ .

<sup>[26]</sup> Para exemplificar, considere-se o caso mais simples em que uma série temporal  $y_t$  ( $t=1, \dots, n$ ) sem sazonalidade está afectada por uma intervenção do tipo "impulso" ocorrida no momento  $t = T$  ( $T < n$ ) e que efeito deste "choque" sobre a série se esgotou nesse mesmo momento do tempo. Nesta situação ter-se-ia



intermédias poderão revelar-se muito mais gravosas em séries com sazonalidade, dado que a exclusão dessas observações poderá distorcer o próprio padrão de sazonalidade da série temporal.

Uma forma de ultrapassar esta situação consiste em excluir, apenas para efeito de cálculo dos coeficientes de autocorrelação, os produtos cruzados que são afectados, nos termos descritos na nota [26]. Para uma série em que foram eliminadas  $(T' - 1 + T_1)$  observações, ter-se-ia:

$$r^*(\tau) = \frac{\sum_{t=1}^{T_1-1-\tau} [(y_t - \hat{\mu}^*)(y_{t+\tau} - \hat{\mu}^*)] + \sum_{t=T'}^{n-\tau} [(y_t - \hat{\mu}^*)(y_{t+\tau} - \hat{\mu}^*)]}{\hat{\sigma}^2(n - T' + T_1 - 2\tau)}$$

$\omega(B) = \omega_0$  e  $\delta(B) = 1$ . Suponha-se, também, para simplificar, que a série não "contaminada" pela intervenção —  $z_t$  — é estacionária de média nula e variância constante  $\sigma^2$ , e que os coeficientes de autocorrelação de  $z_t$  —  $\rho(\tau)$ ,  $(\tau=1,2,\dots)$  — são todos positivos. Neste caso, a estacionaridade desta série garante a verificação da seguinte condição:  $\rho(\tau) \geq \rho(\tau+1)$ .

De acordo com a alternativa ii), a identificação do modelo para  $N_t$  efectuar-se-ia com base nas FAC e FACP de uma "nova" série temporal  $y_t' = \{y_t: t=1,2,\dots,T-1,T+1,\dots,n\}$ . A estimativa do coeficiente de autocorrelação de primeira ordem da série  $y_t'$  viria dado por:

$$\begin{aligned} r^*(1) &= \frac{[(y_2 \cdot y_1) + \dots + (y_{T-1} \cdot y_{T-2}) + (y_{T+1} \cdot y_{T+1}) + (y_{T+2} \cdot y_{T+1}) + \dots + (y_n \cdot y_{n-1})]/(n-2)}{\hat{\sigma}^2} \\ &= \frac{[(z_2 \cdot z_1) + \dots + (z_{T-1} \cdot z_{T-2}) + (z_{T+1} \cdot z_{T+1}) + (z_{T+2} \cdot z_{T+1}) + \dots + (z_n \cdot z_{n-1})]/(n-2)}{\hat{\sigma}^2} \end{aligned}$$

Note-se que o cálculo de  $\hat{\sigma}^2$  não é afectado (a menos do efeito de se dispor de menos uma observação na série colada).

Por sua vez, coeficiente de autocorrelação estimado de ordem um da série não "contaminada" pela intervenção, ou seja,  $z_t$ , vem dado por

$$\begin{aligned} r(1) &= \frac{[(y_2 \cdot y_1) + \dots + (y_{T-1} \cdot y_{T-2}) + [(y_{T-\omega_0} \cdot y_{T-1}) + [y_{T+1} \cdot (y_{T-\omega_0})] + (y_{T+2} \cdot y_{T+1}) + \dots + (y_n \cdot y_{n-1})]/(n-1)]}{\hat{\sigma}^2} \\ &= \frac{[(z_2 \cdot z_1) + \dots + (z_{T-1} \cdot z_{T-2}) + (z_T \cdot z_{T-1}) + (z_{T+1} \cdot z_T) + (z_{T+2} \cdot z_{T+1}) + \dots + (z_n \cdot z_{n-1})]/(n-1)]}{\hat{\sigma}^2} \end{aligned}$$

Comparando as expressões de  $r(1)$  e  $r^*(1)$  pode-se constatar que duas parcelas do coeficiente de autocorrelação estimado de primeira ordem da série  $z_t$  são substituídas, em  $r^*(1)$ , por uma parcela do coeficiente de autocorrelação estimado de segunda ordem. Generalizando para  $s$  desfasamentos ter-se-ia que  $(s+1)$  parcelas de  $r(s)$  são substituídas, em  $r^*(s)$ , por  $s$  parcelas de  $r(s+1)$ . Perante este facto, será de esperar que, em média, os coeficientes de autocorrelação da "nova" série ( $y_t'$ ) venham subestimados, relativamente aos da série não "contaminada".

em que  $T_1$  é o momento de ocorrência da primeira intervenção,  $T^*$  o momento em que já se esgotaram os efeitos da última intervenção,  $\hat{\mu}^*$  e  $\hat{\sigma}^*$  são, respectivamente, a média e a variância amostrais da série temporal  $y_t$ . O valor máximo para  $\tau$  vem condicionado pelo número de observações em cada grupo<sup>[27]</sup>.

iii) admitindo que a variável output é estacionária, identifica-se o modelo ARMA para a componente ruído com base na série:

$$\hat{n}_t = y_t - \hat{c} - \sum_{m=1}^k \hat{v}_m(B) D_{tm} = y_t - \hat{c} - \sum_{m=1}^k \frac{\hat{w}_m(B) B^{b_m}}{\hat{d}_m(B)} D_{tm}$$

sendo  $\hat{n}_t$  uma estimativa de  $N_t$  obtida a partir de uma primeira estimação dos coeficientes dos polinómios  $v_m(B)$  ou  $w_m(B)$  e  $d_m(B)$ , consoante se identifique a função transferência, respectivamente, pelo processo c) ou b) acima indicados. Repare-se que, nesta alternativa, é utilizada a totalidade das observações, procedimento aceitável em situações em que a dimensão da série temporal é razoável e os efeitos das intervenções são meramente temporários.

Quer a alternativa i) quer a alternativa ii), conforme a sugestão apresentada nesta última para o cálculo da fac estimada, têm a vantagem da identificação do modelo ARIMA para a componente residual se efectuar com base em estatísticas — FAC e FACP estimadas — calculadas a partir de dados que, por hipótese, não estarão "contaminados" por "choques" exógenos<sup>[28]</sup>. No entanto, ambas alternativas poderão estar sujeitas a limitações de ordem "quantitativa". Por exemplo, a primeira alternativa pressupõe a existência de um elevado número de observações previamente à ocorrência da primeira intervenção ou após o esgotamento dos efeitos da última, o que implica exigências quanto

---

<sup>[27]</sup> Por exemplo, se  $(T_1 - 1) < (n - T^*)$  e  $\tau \geq (T_1 - 1)$ , então  $r^*(\tau)$  viria dado por

$$r^*(\tau) = \frac{\sum_{t=T^*}^{n-\tau} (y_t - \hat{\mu}^*)(y_{t+\tau} - \hat{\mu}^*)}{\hat{\sigma}^{2*}(n - T^* - \tau)}$$

<sup>[28]</sup> O enviesamento das estatísticas com base nas quais se efectua a identificação do modelo ARIMA (FAC e FACP) constitui uma das consequências da não detecção de "choques" exógenos à série temporal. No capítulo II retomar-se-á este aspecto.

à disponibilidade de informação que só em raras situações se encontrarão satisfeitas. Além disso, existem determinadas séries cronológicas que devido à sua natureza intrínseca estarão afectadas por intervenções em toda a sua extensão. Têm-se como exemplos as séries temporais afectadas pelos efeitos "dias da semana" e "festas móveis"<sup>[29]</sup>.

A alternativa iii) é, por todos os motivos acima referidos, a que é aplicada com maior regularidade em trabalhos empíricos realizados no âmbito dos modelos de análise de intervenção. Uma vez que surge integrada no procedimento iterativo de estimação conjunta dos coeficientes das variáveis de intervenção e do modelo ARIMA da componente residual, efectua-se uma explicação mais pormenorizada desta alternativa nas etapas de estimação e avaliação do diagnóstico.

### 1.2.1.2 - Estimação e Avaliação do Diagnóstico do Modelo de Análise de Intervenção

Após ter-se especificado o modelo

$$y_t = C + \sum_{m=1}^k v_m(B) D_{tm} + N_t = C + \sum_{m=1}^k \frac{\omega_m(B) B^{b_m}}{\delta_m(B)} D_{tm} + N_t \quad (1.3)$$

são dados os seguintes passos:

**1º Passo** - Estimar o modelo (1.3) pelo método da máxima verosimilhança e testar a estacionaridade da primeira estimativa da série ruído ( $\hat{N}_t$ ). No caso de  $N_t$  ser não estacionária os estimadores serão inconsistentes e, como tal, estar-se-á na presença de uma regressão espúria. Em termos gerais, é válido considerar que a ordem de integrabilidade das séries  $y_t$  e  $N_t$  tenderá a ser a mesma, dado que as variáveis exógenas são determinísticas<sup>[30]</sup>.

<sup>[29]</sup> No capítulo III estes dois efeitos serão retomados.

<sup>[30]</sup> A ordem de integrabilidade de uma série pode ser definida como o número mínimo de diferenças de ordem um e  $s$  (sendo  $s$  a periodicidade da série) necessário para tornar a série estacionária. Assim, se  $y_t \sim I(d,D)$  (integrada de ordem  $d$  na componente não sazonal e de ordem  $D$  na componente sazonal), então  $\Delta_s^d \Delta_s^D y_t$  é uma série estacionária, ou seja,  $I(0,0)$ .

Refira-se contudo que, perante a ocorrência de uma quebra de estrutura na componente determinística em que se verifica uma alteração do nível ou da tendência da série, tender-se-á a aceitar mais facilmente a presença de uma raiz unitária na componente não sazonal da série temporal quando, na realidade, ela não existe<sup>[31]</sup>. No entanto, uma vez que as intervenções têm um carácter determinístico, a existência de uma raiz unitária em pelo menos um dos polinómios  $\delta_m(B)$  jamais poderá afectar  $d$  ou  $D$ , ou seja, o grau de integrabilidade da série, atendendo a que este está relacionado com a componente estocástica da mesma.

Feita esta ressalva, pode então aplicar-se o mesmo número de diferenças, de ordem  $um$  e  $s$ , que tornam a série  $y_t$  estacionária à série ruído.

$$\Delta^d \Delta_s^D N_t = (1-B)^d (1-B^s)^D N_t = \zeta_t \quad (1.4)$$

Inserindo (1.4) em (1.3), obtém-se o seguinte modelo:

$$\Delta^d \Delta_s^D y_t = \Delta^d \Delta_s^D C + \sum_{m=1}^k \frac{\omega_m(B) B^{b_m}}{\delta_m(B)} \Delta^d \Delta_s^D D_{tm} + \zeta_t \quad (1.5)$$

onde a primeira parcela do lado direito será nula se  $d \geq 1$  ou  $D \geq 1$ .

Existem autores (e.g., Liu (1986)) que aconselham que se especifique, logo à partida, um modelo ARMA  $(1,0)(1,0)_s$  para a série residual do modelo (1.3)<sup>[32]</sup>. Esta especificação, ao contemplar um conjunto de casos particulares relevantes, acaba por se revelar bastante vantajosa. Por exemplo, a obtenção de estimativas dos parâmetros auto-regressivos próximas da unidade indicia a existência de raízes unitárias nas componentes sazonal e não sazonal sendo, portanto, necessárias uma diferença de ordem  $um$  e uma de ordem para estacionarizar a série. Estes autores afirmam que com este procedimento menor é o risco de se incorrer numa sobre-diferenciação da série temporal. Se, por outro lado, as estimativas forem estatisticamente não significativas, então maior é a garantia

---

<sup>[31]</sup> Refira-se aqui a proposta de teste de raízes unitárias desenvolvida por Perron (1988) face à inadequação dos testes "tradicionais" de raízes unitárias (Dickey-Fuller — DF — e Dickey-Fuller aumentado — ADF) em situações de quebra de estrutura na parte determinística.

<sup>[32]</sup> Para séries sem sazonalidade propõem que se especifique um modelo AR(1) ou AR(2).

de que a série residual não se afasta muito da série ruído branco. Outra das vantagens desta abordagem é a de que, com base nesta especificação inicial, a tarefa de identificação do modelo para a série residual, a realizar numa fase posterior, fica muito mais facilitada.

No modelo (1.5) está assegurada a consistência dos estimadores dos parâmetros mas, se os  $\zeta_t$ 's forem autocorrelacionados, i.e., se a série  $\{\zeta_t\}$  não for um processo ruído branco, eles não serão os mais eficientes.

**2º Passo** - Para melhorar a eficiência dos estimadores dos coeficientes é aconselhável especificar para a variável residual um modelo ARMA que represente uma boa aproximação ao verdadeiro processo de  $\zeta_t$ . Para tal, pode-se utilizar como referência a seguinte série estimada:

$$\hat{\zeta}_t = \Delta^d \Delta_s^D y_t - \Delta^d \Delta_s^D \hat{c} - \sum_{m=1}^k \Delta^d \Delta_s^D \frac{\hat{\omega}_m(B) B^{b_m}}{\hat{\delta}_m(B)} D_{tm}$$

em que  $\hat{c}$ ,  $\hat{\omega}_m(B)$  e  $\hat{\delta}_m(B)$  são, respectivamente, as estimativas iniciais de  $C$ ,  $\omega_m(B)$  e  $\delta_m(B)$ , obtidas no primeiro passo por estimação de (1.5).

**3º Passo** - Re-estimação do modelo (1.5), após se ter substituído  $\zeta_t$  pelo modelo ARMA obtido no passo 2, ou seja, estimação do modelo (1.6):

$$\Delta^d \Delta_s^D y_t = \Delta^d \Delta_s^D C + \sum_{m=1}^k \frac{\omega_m(B) B^{b_m}}{\delta_m(B)} \Delta^d \Delta_s^D D_{tm} + \frac{\Theta_q(B^*) \theta_q(B)}{\Phi_p(B^*) \phi_p(B)} \varepsilon_t \quad (1.6)$$

A estimação conjunta dos coeficientes das variáveis de intervenção e dos parâmetros do modelo ARIMA para a componente ruído permite obter ganhos de eficiência na fase de estimação. A especificação "final" do modelo é obtida ao fim de um processo iterativo

tendente à obtenção de um modelo mais parcimonioso em que, tendo sido realizados os testes do controlo do diagnóstico<sup>[33]</sup>, não foram detectadas quaisquer incorrecções.

Em termos conclusivos pode-se afirmar que, em séries temporais afectadas por intervenções exógenas, e supondo correctas as especificações propostas para modelar os seus efeitos, a metodologia da análise de intervenção tende a ser claramente superior aos modelos ARIMA univariados no que respeita à qualidade do ajustamento que se obtém. Todavia, é de referir que ao nível da previsão o mesmo pode não acontecer. Isto porque, as variáveis de intervenção, ao representarem "choques" exógenos cujo momento de ocorrência é conhecido no período amostral, são imprevisíveis em termos futuros e, como tal, os "ganhos" em termos previsionais dos modelos de intervenção relativamente aos ARIMA univariados poder-se-ão revelar pouco significativos.

---

<sup>[33]</sup> Note-se que, exceptuando a componente que respeita às variáveis de intervenção, a avaliação do diagnóstico do modelo de intervenção processa-se nos mesmos moldes que o controlo de diagnóstico de um modelo ARIMA univariado. Para além de verificar a significância estatística das estimativas dos parâmetros do modelo da componente  $N_t$ , são realizados os habituais testes sobre os resíduos do modelo (1.6): os testes de Bartlett e de Daniels-Jenkins sobre, respectivamente, os coeficientes de autocorrelação e de autocorrelação parcial da série residual e o teste Ljung-Box que admite como hipótese nula o "comportamento" ruído branco da mesma.

## II - OUTLIERS EM SÉRIES TEMPORAIS: Procedimento Iterativo de Detecção e Correção

No capítulo precedente apresentou-se a metodologia que deve estar subjacente ao tratamento de séries temporais afectadas por fenómenos exógenos e determinísticos cujos momentos de ocorrência no tempo são conhecidos *a priori*. Existem, contudo, situações em que, em virtude de se desconhecer a ocorrência desses fenómenos, se acaba por ignorar os seus efeitos sobre série temporal em estudo. Atendendo às consequências nefastas que este procedimento tende a ter ao nível das etapas de identificação, estimação e previsão do processo de modelação univariado, tornou-se premente o desenvolvimento de técnicas que permitissem alcançar um maior grau de fiabilidade nos resultados obtidos da especificação e estimação de modelos ARIMA para séries temporais afectadas por observações atípicas.

Por conseguinte, surgiram na literatura várias abordagens com o objectivo de incorporar na modelação Box-Jenkins univariada a detecção e correção de observações "anormais" designadas, em terminologia inglesa, por *outliers*. Entre elas destacam-se a abordagem bayesiana proposta por Abraham e Box (1979), os métodos robustos aconselhados por Denby e Martin (1979) e um conjunto de procedimentos iterativos desenvolvidos por Chang (1982), Chang e Tiao (1983), Tsay (1986), Chang, Tiao e Chen (1988), Tsay (1988), Chen e Liu (1990) e Chen e Liu (1993a). A simplicidade inerente a estes últimos e a facilidade de implementação a nível computacional levou a que, em aplicações empíricas, os procedimentos iterativos assumissem um papel de destaque na análise de *outliers*, relativamente às outras abordagens. Tendo presente este aspecto, far-se-á neste capítulo uma breve descrição do procedimento proposto por Chen e Liu atendendo a que é considerado, actualmente, um dos procedimentos mais recentes para detectar e corrigir *outliers* e, mais importante, porque está incorporado no módulo das séries temporais do programa SCA (*Scientific Computing Associates*), *software* que se utilizou na aplicação empírica realizada neste trabalho.

## II.1 - Os Diversos Tipos de *Outliers*

Os *outliers*, no contexto das séries temporais, podem ser definidos como observações que parecem ser totalmente inconsistentes com o resto da amostra. Embora a sua origem seja à partida desconhecida, alguns autores (e.g., Tsay (1986)) aconselham que, uma vez detectados, se indague sobre as possíveis causas da sua ocorrência de modo a incorporá-los no modelo como variáveis de intervenção. Na literatura são habitualmente considerados quatro tipos de *outliers*, o AO (*additive*), o IO (*innovational*), o LS (*level step*) e o TC (*temporary change*), embora os artigos pioneiros nesta área, entre eles o de Fox (1972), tenham apenas considerado os dois primeiros tipos. Tsay (1988), por exemplo, designa por *outliers* os do tipo AO e IO, mas considera os do tipo LS e TC como fenómenos de quebra de estrutura.

Em termos simplificados, podemos dizer que a série temporal está "contaminada" por um AO quando se regista um "salto" na série em apenas uma observação, que respeita ao momento de ocorrência do "choque"<sup>[1]</sup>, enquanto que se se verificar uma alteração do nível da mesma a partir desse momento diz-se, então, que o *outlier* é do tipo LS. No caso do *outlier* TC, os seus efeitos tendem a decrescer exponencialmente no tempo, sendo eliminados ao fim de um número finito de observações. O IO, como o seu nome indica, corresponde a um "choque" na série residual, também designada por série das inovações, cujos efeitos se propagam no tempo de acordo com a estrutura do modelo subjacente à série temporal.

Desta forma, designando por  $y_t$  a série que está afectada por um *outlier* na observação  $t=T$ , e sobre a qual existem observações disponíveis, e por  $z_t$  a série não "contaminada" por *outliers* associada a  $y_t$ , tem-se a seguinte relação:

$$y_t = f(t) + z_t \quad (2.1)$$

em que  $f(t)$  é uma função paramétrica determinística que representa os efeitos dos "choques" exógenos sobre a série  $z_t$  não observável.

---

<sup>[1]</sup> Em inglês, é usual designar-se este *outlier* por *gross error*, atendendo a que tende a estar muitas vezes associado a erros de medida.



Usualmente admite-se para  $f(t)$  uma especificação do tipo

$$f(t) = \omega_0 \frac{\omega(B)}{\delta(B)} D_t \quad (2.2)$$

onde  $D_t$  é uma variável binária que assume o valor 1 em  $t=T$  e o valor 0 nos restantes casos. Repare-se que  $(\omega_0 \omega(B)/\delta(B))$  pode ser vista como uma função transferência, semelhante às apresentadas no primeiro capítulo, em que  $\omega_0$  mede o efeito imediato e  $\omega(B)/\delta(B)$  traduz os efeitos dinâmicos decorrentes do "choque" exógeno.

Suponha-se, agora, que a série  $z_t$  segue um processo ARIMA(p,d,q) do tipo  $\phi_p(B)(1-B)^d z_t = \theta_q(B) \varepsilon_t$  [2], e que

$$\frac{\phi_p(B) (1-B)^d}{\theta_q(B)} = \pi(B) = (1 - \pi_1 B - \pi_2 B^2 - \pi_3 B^3 - \dots) \quad (2.3)$$

Tendo em conta as especificações (2.1) e (2.2) e a forma como acima se definiu os diferentes tipos de *outliers*, tem-se a seguinte função transferência associada a cada um deles

$$\left\{ \begin{array}{l} \text{A0: } \omega_0 \frac{\omega(B)}{\delta(B)} = \omega_A \end{array} \right. \quad (2.4)$$

$$\left\{ \begin{array}{l} \text{LS: } \omega_0 \frac{\omega(B)}{\delta(B)} = \frac{\omega_{LS}}{1-B} \end{array} \right. \quad (2.5)$$

$$\left\{ \begin{array}{l} \text{TC: } \omega_0 \frac{\omega(B)}{\delta(B)} = \frac{\omega_{TC}}{1-\delta B} \quad (0 < \delta < 1) \end{array} \right. \quad (2.6)$$

$$\left\{ \begin{array}{l} \text{IO: } \omega_0 \frac{\omega(B)}{\delta(B)} = \omega_{IO} \pi(B)^{-1} = \omega_{IO} \frac{\theta_q(B)}{(1-B)^d \phi_p(B)} \end{array} \right. \quad (2.7)$$

---

[2] Assume-se aqui, para facilitar a análise, que a série  $z_t$  é uma série sem componente sazonal, embora o procedimento que se apresenta seja aplicável a séries temporais com sazonalidade.

Repare-se que, ao contrário dos *outliers* do tipo AO, LS e TC, a especificação para o IO depende da estrutura do modelo subjacente à série  $z_t$ . Por outro lado, é fácil verificar que os *outliers* do tipo AO e LS podem ser vistos como casos extremos do *outlier* do tipo TC em que, no primeiro caso,  $\delta=0$  e, no segundo,  $\delta=1$ . O parâmetro  $\delta$ , que marca o ritmo a que o efeito gerado pelo "choque" exógeno se amortece ao longo do tempo, é, na prática, fixado à *priori*. Chen e Liu (1993a), por exemplo, propõem  $\delta=0.7$  considerando este valor como uma boa solução de compromisso para distinguir este tipo de *outlier* dos do tipo AO e LS, enquanto que Tsay (1988) considera um valor de 0.8.

Refira-se, ainda, que existem alguns autores, e.g. Tsay (1988), que além dos quatro tipos de *outliers* atrás definidos consideram um outro tipo de *outlier* que designam por VC (*variance change*). Este tipo de "choque" exógeno tem a particularidade de afectar não só a média da série  $z_t$  mas, também, a variância da mesma. Atendendo a que o programa SCA não tem incorporado o procedimento de detecção e ajustamento do *outlier* VC, ele não será objecto de estudo neste trabalho<sup>[3]</sup>.

## II.2 - Consequências da Presença de *Outliers* nos Resultados de Modelação e Previsão Box-Jenkins Univariada

Estudos efectuados sobre os efeitos de observações atípicas nos resultados das etapas de identificação, estimação e previsão do procedimento Box-Jenkins univariado<sup>[4]</sup> concluíram que a existência destas observações atípicas pode causar um enviesamento substancial nos estimadores dos coeficientes de autocorrelação e autocorrelação parcial pondo, desta forma, em causa a utilização destas estatísticas como instrumentos de identificação do modelo ARIMA. Consequentemente, maior é o risco de se incorrer numa má especificação do modelo e, como tal, obter uma especificação final que poderá estar sub ou sobre-parametrizada. Mesmo que o modelo esteja bem especificado, os estimadores dos respectivos parâmetros tenderão a vir enviesados, pelo que qualquer inferência que se realize sobre os mesmos poderá ser pouco credível, além de que, e como seria de esperar, o não ajustamento de *outliers* durante o processo de estimação tende a inflacionar a

---

<sup>[3]</sup> Refira-se, contudo, que Tsay (1988) apresenta um procedimento iterativo de detecção deste tipo de *outlier*.

<sup>[4]</sup> Tsay (1986) aponta os de Guttman e Tiao (1978), Miller (1980) e Chang (1982).

estimativa da variância da variável residual. Note-se, contudo, que o grau de enviesamento introduzido pela presença de observações "anormais" na série temporal depende do número, tipo, magnitude e posição relativa dos *outliers* na respectiva série temporal.

Ao nível da previsão, a presença de *outliers* na série temporal produz também enviesamentos nos resultados. Ledolter (1989) estudou os efeitos de um *outlier* aditivo sobre as previsões dos modelos ARIMA e constatou um aumento do erro quadrático médio de previsão, resultante quer da transmissão dos efeitos do AO ao período de previsão quer do enviesamento que se tende a registar nos estimadores dos parâmetros do modelo ARIMA devido à ocorrência desse "choque" exógeno. Concluiu, ainda, que esse aumento é tanto maior quanto mais o próximo estiver o momento de ocorrência do AO da origem da previsão, pelo que, se o "choque" ocorrer na origem da previsão os seus efeitos sobre a qualidade da previsão tenderão a ser desastrosos. Ao nível dos intervalos de previsão, verificou que estes tendem a ser bastantes afectados, dado que, e como atrás se referiu, os efeitos do *outlier* tendem a inflacionar a variância do erro da previsão e, consequentemente, a aumentar a amplitude do intervalo de previsão. Chen e Liu (1993b), num estudo similar ao de Ledolter (1989) em que estendem a análise aos outros tipos de *outliers* (LS, TC e IO), chegam às mesmas conclusões que as de Ledolter e aconselham que, para efeitos de previsão, seja considerado, no procedimento de detecção de *outliers* aplicado às observações localizadas perto da origem de previsão, um valor crítico menor.

### **II.3 - Estimação dos Efeitos dos *Outliers* Assumindo que os Parâmetros do Modelo ARIMA são Conhecidos**

As observações atípicas podem ser vistas como o resultado de "choques" exógenos não antecipados ou, mais concretamente, como "choques" cujo momento de ocorrência é desconhecido à partida, pelo que a detecção de *outliers* tende a ser efectuada a partir da análise dos resíduos estimados do modelo ARIMA especificado para a série temporal em estudo.

Tendo em vista estimar os efeitos de cada um dos *outliers* e testar a sua existência no momento  $t=T$  assume-se, por agora, que os parâmetros do modelo ARIMA( $p,d,q$ ) e a

variância da série residual  $\{\varepsilon_t\}$ ,  $\sigma_\varepsilon^2$ , são conhecidos. Como tal, a série residual, eventualmente "contaminada" por *outliers*, pode ser expressa da seguinte forma

$$e_t = \pi(B) y_t \quad t=1, \dots, n \quad (2.8)$$

Substituindo (2.1) em (2.8) e considerando as especificações (2.4) a (2.7) tem-se o seguinte modelo de regressão para cada um dos tipos de *outliers* atrás referidos:

$$\left\{ \begin{array}{l} \text{AO: } e_t = \omega_A \pi(B) D_t + \varepsilon_t \end{array} \right. \quad (2.9)$$

$$\left\{ \begin{array}{l} \text{LS: } e_t = \omega_{LS} \frac{\pi(B)}{1-B} D_t + \varepsilon_t \end{array} \right. \quad (2.10)$$

$$\left\{ \begin{array}{l} \text{TC: } e_t = \omega_{TC} \frac{\pi(B)}{1-\delta B} D_t + \varepsilon_t \end{array} \right. \quad (2.11)$$

$$\left\{ \begin{array}{l} \text{IO: } e_t = \omega_{IO} D_t + \varepsilon_t \end{array} \right. \quad (2.12)$$

Atendendo a que todos os parâmetros das regressões (2.9) a (2.12) são conhecidos, à excepção do parâmetro  $\omega_i$  (para  $i = \text{AO, LS, TC e IO}$ ), estas podem ser re-escritas do seguinte modo

$$e_t = \omega_i x_{it} + \varepsilon_t \quad t = T, T+1, \dots, n \quad (2.13)$$

onde

$$\left\{ \begin{array}{l} x_{it} = 0, \quad \text{para } t < T, \forall i \\ x_{it} = 1, \quad \text{para } t = T, \forall i \\ x_{\text{AO}, T+k} = -\pi_k \\ x_{\text{LS}, T+k} = 1 - \sum_{j=1}^k \pi_j \\ x_{\text{TC}, T+k} = \delta^k - \sum_{j=1}^k \delta^{k-j} \pi_j \\ x_{\text{IO}, T+k} = 0 \end{array} \right\} \quad \text{para } k \geq 1$$

Da aplicação do método dos mínimos quadrados à regressão (2.13), a estimativa do parâmetro  $\omega$  associada a cada um dos tipos de *outliers* vem dada por

$$\hat{\omega}_i = \frac{\sum_{t=T}^n e_t x_{it}}{\sum_{t=T}^n x_{it}^2} \quad \text{para } i=AO,LS,TC,IO \quad (2.14)$$

Repare-se que no caso do IO a expressão (2.14) reduz-se a  $\hat{\omega}_{IO} = e_T$ , ou seja, o melhor estimador do efeito do IO que ocorre no momento  $t=T$  é o resíduo  $e_T$ . Por outro lado, note-se que se o *outlier* ocorrer na última observação da amostra, ou seja, em  $T=n$ , então  $\hat{\omega}_{AO}=\hat{\omega}_{LS}=\hat{\omega}_{TC}=\hat{\omega}_{IO}=e_n$ , não sendo, portanto, possível distinguir o tipo de *outlier*.

Uma possível abordagem para testar a existência de um *outlier* do tipo  $i$ , no momento  $t=T$ , foi proposta por Chang, Tiao e Chen (1988), na qual se examina o valor máximo de um conjunto de quatro estatísticas estandardizadas —  $\tau_i(T)$  — que correspondem, ao fim e ao cabo, à habitual estatística teste da nulidade do parâmetro numa regressão linear, calculada para cada um dos tipos de *outliers*. Desta forma, tem-se

$$\tau_i(T) = \frac{\hat{\omega}_i}{\sigma_{\hat{\omega}_i}} = \frac{\hat{\omega}_i}{\sigma_e \sqrt{\sum_{t=T}^n x_{i,t}^2}} \quad i=AO,LS,TC,IO \quad (2.15)$$

Sob a hipótese nula de que não existe um *outlier* na série  $z_t$  no momento  $t=T$ , contra a alternativa da existência de um *outlier* do tipo  $i$ , a estatística  $\tau_i(T)$  tem uma distribuição  $N(0,1)$ . Se os parâmetros do modelo ARIMA e o desvio padrão da série residual não forem conhecidos (e é o que acontece na prática), mas forem substituídos por estimadores consistentes, então estas estatísticas terão uma distribuição assintótica  $N(0,1)$ .

Deste modo, após se terem calculado as quatro estatísticas de teste, é detectado um *outlier* do tipo  $i$  no momento  $t=T$  quando o máximo do conjunto  $\{|\tau_i(T)|, i=AO,LS,TC,IO\}$  for superior a um determinado valor crítico positivo pré-especificado. Com base em estudos, em que utilizam a simulação, Chang, Tiao e Chen (1988) aconselham que se

considerem valores críticos entre 3.0 e 4.0, enquanto que Chen e Liu (1993a) propõem valores inferiores, igual ou superiores a 3.0, consoante a série tenha um número de observações, respectivamente, inferior a 150, entre 150 e 250, ou superior a 250. A série ajustada dos efeitos do *outlier* pode ser calculada com base nas expressões (2.1) e (2.2), em que se substitui o parâmetro  $\omega_0$  da função  $f(t)$  pela estimativa  $\hat{\omega}_i$  obtida na expressão (2.14).

#### II.4 - Procedimento de Detecção e Estimação dos Efeitos de Múltiplos *Outliers* em Séries Temporais quando os Parâmetros do Modelo ARIMA são Desconhecidos

Na metodologia atrás descrita, para detectar o tipo e efeitos de um *outlier* que ocorre no momento  $t=T$ , assumiu-se que os parâmetros do modelo ARIMA da série  $z_t$  são conhecidos. Todavia, na prática, eles são desconhecidos, pelo que se torna necessário estimá-los previamente à detecção dos *outliers*. Por outro lado, esse procedimento revela-se insuficiente em situações em que a série temporal está "contaminada" por múltiplos *outliers* cujos momentos de ocorrência são desconhecidos. Na realidade, no caso em que apenas existe um *outlier*, mesmo que se desconheça o momento em que ocorre, é relativamente simples determinar a sua localização bem como o seu tipo. Por exemplo, é detectado um *outlier* do tipo  $i$  no momento  $T'$  ( $1 \leq T' \leq n$ ) quando

$$\tau_i(T') = \max_{(i,t)} \{ |\tau_i(t)|, i = AO, LS, TC, IO \text{ e } t = 1, \dots, n \} > C \quad (2.16)$$

em que  $C$  é um valor crítico pré-fixado.<sup>[6]</sup>

Todavia, caso a série esteja "contaminada" por vários *outliers*, este procedimento poder-se-á revelar inadequado pois, não só se torna complicado estimar os seus efeitos numa base individual, devido à potencial sobreposição de efeitos, como, também, maior é o perigo de se incorrer em detecções espúrias.

---

<sup>[6]</sup> Repare-se que o cálculo de  $\tau_i(T')$  dado por (2.16) implica a estimação de  $n \times 4$  regressões ( $n$  por cada um dos tipos de *outliers*).

Tendo em vista solucionar estes problemas, vários autores (e.g., Chang, Tiao, Tsay, Chen e Liu) propuseram procedimentos iterativos que asseguram, com um razoável grau de fiabilidade, a correcta detecção e estimação do número, tipo e magnitude dos *outliers* e, conseqüentemente, a identificação de um modelo ARIMA apropriado para séries temporais afectadas por observações atípicas. Entre eles, destaca-se o procedimento desenvolvido por Chen e Liu (1993a) que, segundo os autores, tem como vantagens permitir que os efeitos e tipos de *outliers* sejam estimados com base em estimativas dos parâmetros do modelo ARIMA menos "contaminadas", evitar a detecção de *outliers* espúrios e, ainda, estimar conjuntamente os parâmetros do modelo e os efeitos dos *outliers*. Como se referiu anteriormente, o procedimento desenvolvido por Chen e Liu está incorporado no SCA, programa que se utilizou na aplicação empírica que se apresenta no capítulo IV. Atendendo a que o SCA detecta e ajusta os *outliers* de uma forma automática com uma simples instrução, parece conveniente explicar, nesta secção, ainda que de forma muito breve, o procedimento que lhe está subjacente<sup>[6]</sup>.

Suponha-se a situação em que a série  $y_t$  está "contaminada" por  $k$  *outliers*, cujos momentos de ocorrência são desconhecidos. Tem-se, então, o seguinte modelo

$$y_t = \sum_{m=1}^k \omega_{0m} \frac{\omega_m(B)}{\delta_m(B)} D_t(T_m) + z_t = \sum_{m=1}^k \omega_{0m} \frac{\omega_m(B)}{\delta_m(B)} D_t(T_m) + \frac{\theta_q(B)}{(1-B)^d \phi_p(B)} \epsilon_t \quad (2.17)$$

onde  $t_m$  é o momento de ocorrência do  $m$ -ésimo "choque" exógeno e  $(\omega_{0m} \omega_m(B)/\delta_m(B))$  a respectiva função transferência, cujas especificações, para cada tipo de *outlier*, estão indicadas nas expressões (2.4) a (2.7). Neste caso, os parâmetros de  $\theta_q(B)$  e  $\phi_p(B)$  são, à partida, desconhecidos.

O procedimento proposto por Chen e Liu engloba três grandes etapas que se desenvolvem segundo uma lógica sequencial-iterativa.

Na primeira etapa, procede-se à estimação preliminar dos parâmetros do modelo ARIMA, especificado para a série original ( $y_t$ ), e à detecção dos *outliers*. Começa-se por

---

<sup>[6]</sup> Uma descrição mais pormenorizada do mesmo pode ser encontrada em Chen e Liu(1990) e Chen e Liu (1993a).

estimar o modelo (2.17), ignorando a componente sistemática respeitante aos efeitos das variáveis  $D_m$  e, com base na estimativa da respectiva série residual, aplica-se o procedimento de detecção de *outliers* descrito na secção II.3, mas levando agora em conta que o momento de ocorrência é desconhecido<sup>[7]</sup>. Deste modo, em cada iteração é apenas detectado um *outlier* cuja localização e tipo são determinados pela condição (2.16). Após a correcção da estimativa da série residual dos efeitos deste *outlier*, recorrendo às expressões (2.9) a (2.12), torna-se a aplicar o procedimento de detecção de *outliers* (sem re-estimação do modelo). Quando mais nenhum *outlier* for detectado, ajusta-se a série  $y_t$  dos efeitos de todos os *outliers* encontrados, recorrendo às expressões (2.4) a (2.7), e especifica-se e estima-se um modelo ARIMA para a série ajustada. O enviesamento dos estimadores dos parâmetros deste modelo tenderá, por esta razão, a ser menor. Com base na estimativa da série residual deste "novo" modelo procede-se novamente à detecção dos *outliers*. Este procedimento de detecção e re-estimação dos parâmetros do modelo ARIMA é repetido até que mais nenhum *outlier* seja detectado nesta primeira etapa.

Na segunda etapa, estima-se conjuntamente os efeitos dos *outliers* e os parâmetros do modelo, existindo a preocupação em eliminar quaisquer efeitos espúrios que resultem do procedimento iterativo descrito na primeira etapa. Para tal é especificado e estimado um modelo de regressão linear múltipla do tipo

$$e_t = \sum_{m=1}^k \omega_{0m} \frac{\omega_m(B)}{\delta_m(B)} D_t(T_m) + u_t \quad (2.18)$$

onde  $\{e_t\}$  é a estimativa da série residual obtida na última re-estimação do modelo ARIMA efectuada na primeira etapa e  $\omega_m(B)/\delta_m(B) D_t(T_m)$  são as variáveis exógenas<sup>[8]</sup>. Estimado o modelo (2.18), calcula-se a estatística  $\tau_m = \hat{\omega}_{0m} / \sigma_{\hat{\omega}_{0m}}$ , para  $m=1, \dots, k$ , e determina-se o  $\min_m |\tau_m|$  que, para simplificar a notação, se designa por  $\tau_{\min}$ . O *outlier* associado à estatística de teste  $\tau_{\min}$  é considerado espúrio se  $\tau_{\min}$  for inferior a um determinado valor

---

<sup>[7]</sup> Repare-se, contudo, que, se existirem observações atípicas na série temporal, os estimadores do modelo ARIMA, obtidos nesta estimação preliminar, tenderão a ser enviesados.

<sup>[8]</sup> Note-se que o número de variáveis exógenas vai ser igual ao número total de *outliers* detectados na primeira etapa.



crítico<sup>[9]</sup> e, nesse caso, é eliminado do modelo de regressão. Considerando agora os restantes  $k-1$  *outliers*, torna-se a estimar a regressão (2.18) e a testar a existência de mais algum *outlier* espúrio. Este processo termina quando todas as estimativas dos coeficientes do modelo (2.18) forem estatisticamente significativas. Com base nessas estimativas ajusta-se a série original dos efeitos dos *outliers* e, em seguida, especifica-se e estima-se um modelo ARIMA para a série ajustada. Consideram-se as estimativas dos parâmetros deste modelo como estimativas finais.

A terceira etapa constitui, de certa forma, uma repetição dos passos dados na primeira etapa, só que, neste caso, a detecção dos *outliers* é realizada tomando como referência a estimativa da série residual da especificação "final" do modelo ARIMA, obtida no final da segunda etapa, cujos coeficientes estimados estarão quase totalmente isentos de enviesamentos devidos aos efeitos dos *outliers*. O procedimento termina quando mais nenhum *outlier* for detectado.

---

<sup>[9]</sup> Toma-se como valor crítico o mesmo valor que se considerou na etapa de detecção dos *outliers* (veja-se a secção II.3).

### III - EFEITO "DIAS DA SEMANA" E EFEITO "PÁSCOA": Procedimentos de Detecção e Correção

No capítulo I foi referido que a metodologia da análise de intervenção pode ser considerada como um caso particular dos modelos de função transferência (MFT), embora alguns autores prefiram considerá-la como uma generalização da análise com variáveis artificiais (*dummies*). A primeira ideia parece, contudo, ser mais adequada, uma vez que os procedimentos subjacentes aos MFT são também apropriados aos modelos de análise de intervenção<sup>[1]</sup> e, por outro lado, porque as variáveis de intervenção podem assumir uma estrutura distinta das variáveis binárias. Têm-se como exemplos as variáveis de intervenção que visam captar os efeitos de "calendário". Embora não seja plenamente aceitável considerar este tipo de variáveis como variáveis de intervenção "típicas"<sup>[2]</sup>, é razoável integrá-las nos modelos de análise de intervenção e dar-lhes um tratamento idêntico às variáveis de intervenção "propriamente ditas", dado que a sua natureza é semelhante à destas últimas.

Os efeitos de "calendário", aqui divididos em efeito "dias da semana" e efeito "Páscoa", estão presentes em séries temporais cujo nível, num determinado período de tempo, por exemplo, mês ou trimestre, é afectado por variações na composição do calendário. Consequentemente, este tipo de variações, ao introduzirem alguma variabilidade adicional nos dados da série, constitui um factor que poderá dificultar a análise do comportamento da variável em questão, sendo, deste modo, aconselhável a sua eliminação. Existem na literatura diversos procedimentos para detectar e corrigir os efeitos de "calendário", alguns dos quais serão abordados neste capítulo. Começa-se por abordar os relativos ao efeito "dias da semana", seguindo-se os referentes ao efeito "Páscoa" ou, mais genericamente, efeito "festas móveis".

---

<sup>[1]</sup> Com excepção dos procedimentos utilizados na fase de identificação dos MFT, dado que, e como foi mencionado anteriormente, as variáveis de intervenção assumem um carácter determinístico.

<sup>[2]</sup> Entenda-se variáveis de intervenção "típicas" como as que são introduzidas no modelo ARIMA univariado com o objectivo de quantificar os efeitos de um "choque" exógeno que ocorre num determinado momento do tempo. Em geral, este tipo de "choques" assumem um carácter meramente pontual ou, pelo menos, são fenómenos que não se repetem no tempo com uma determinada periodicidade.

### **III.1 - Efeito "Dias da Semana"**

O efeito "dias da semana" está presente numa série cronológica quando, tomando como referência a unidade de tempo o "dia", o nível da série depende do dia da semana em que ele está a ser medido. Consequentemente, o valor do agregado mensal<sup>[3]</sup> irá depender do número de vezes que cada dia da semana ocorre no mês em questão. Como a composição do calendário se altera de mês para mês e, para o mesmo mês, ao longo dos anos, pode acontecer que do cálculo das taxas de variação homólogas e das taxas em cadeia resultem valores "anormais", pela simples razão de, por exemplo, o número de segundas-feiras, sextas-feiras ou domingos ter sido diferente nos respectivos meses.

A existência de séries temporais afectadas por este tipo de variações na composição do calendário pode ficar a dever-se basicamente a duas ordens de motivos:

i) à própria natureza da série temporal. Esta constitui, sem dúvida, a principal justificação da existência deste tipo de fenómeno. Como exemplos de séries incluídas neste grupo têm-se: o índice de produção industrial (IPI), as vendas a retalho, etc.

ii) a outras causas "exteriores", como sejam, determinadas práticas de contabilização dos dados. Por exemplo, uma unidade que efectue a sua contabilidade numa base semanal, num determinado dia da semana (e.g., todas as sextas-feiras), poderá registar um acréscimo de actividade que se deve, pura e simplesmente, à existência de uma sexta-feira adicional no mês em causa.

#### **III.1.1 - Correção do Efeito "Dias da Semana"**

Os procedimentos existentes para corrigir o efeito "dias da semana" consistem, basicamente, em extensões de metodologias já estabelecidas. Por exemplo, e como à frente

---

<sup>[3]</sup> Refira-se que o efeito "dias da semana" tenderá a ocorrer com maior frequência em séries de periodicidade mensal, ao invés das de periodicidade trimestral, dado que as variações devidas a este tipo de fenómeno tendem a anular-se em 2 ou 3 meses. Como tal, o efeito "dias de semana" ter-se-á praticamente esgotado ao fim de um trimestre.

se desenvolverá, Hillmer (1982) propõe a incorporação de variáveis "dias da semana"<sup>(4)</sup> nos modelos ARIMA univariados, enquanto que Harvey (1990), no contexto dos modelos estruturais, introduz uma componente adicional para captar o referido efeito.

A abordagem tradicional, e a mais frequentemente aplicada, para corrigir o efeito "dias da semana" compreende duas formas alternativas de modelação devidas a Hillmer, Bell e Tiao (1981), Hillmer (1982), Cleveland e Grupe (1982) que, como adiante se demonstrará, estão inter-relacionadas, podendo ser consideradas substitutas entre si.

A primeira alternativa consiste em incluir no modelo, por exemplo, no modelo ARIMA univariado, sete variáveis  $Z_{jt}$  ( $j=1,2,...,7$ ), em que  $Z_{jt}$  mede o número de dias do "tipo"  $j$  que o mês  $t$  contém, isto é, o número de segundas-feiras, terças-feiras, etc.. Nesta alternativa, a componente que quantifica o efeito "dias da semana" ( $DS_t$ ), no mês  $t$ , virá dada por:

$$DS_t = \sum_{j=1}^7 \alpha_j Z_{jt} \quad (3.1)$$

em que  $\alpha_j$  é o coeficiente associado ao dia  $j$ .

Uma segunda alternativa consiste em especificar  $DS_t$  na forma

$$DS_t = \sum_{j=1}^7 \beta_j T_{jt} = \sum_{j=1}^6 \beta_j (Z_{jt} - Z_{7t}) + \beta_7 \sum_{j=1}^7 Z_{jt} \quad (3.2)$$

onde  $T_{jt}$  ( $j=1,...,6$ ) mede a diferença entre o número de dias do tipo  $j$  e o número de domingos no mês  $t$  e  $T_{7t}$  mede o número total de dias do mês  $t$ .

---

<sup>(4)</sup> Em inglês, denominadas variáveis *trading day*.

Atendendo à definição das variáveis  $Z_{jt}$  e  $T_{jt}$  tem-se a seguinte relação entre os parâmetros das especificações (3.1) e (3.2):

$$\begin{aligned}
 DS_t &= \sum_{j=1}^7 \beta_j T_{jt} = \sum_{j=1}^6 \beta_j T_{jt} + \beta_7 T_{7t} = \sum_{j=1}^6 \beta_j (Z_{jt} - Z_{7t}) + \beta_7 \sum_{j=1}^7 Z_{jt} \\
 &= \sum_{j=1}^6 \beta_j Z_{jt} - Z_{7t} \sum_{j=1}^6 \beta_j + \beta_7 \sum_{j=1}^7 Z_{jt} = \sum_{j=1}^6 (\beta_j + \beta_7) Z_{jt} + (\beta_7 - \sum_{j=1}^6 \beta_j) Z_{7t} \\
 &= \sum_{j=1}^7 \alpha_j Z_{jt}
 \end{aligned}$$

Resolvendo o sistema

$$\begin{cases} \alpha_j = \beta_j + \beta_7, & \text{para } j=1, \dots, 6 \\ \alpha_7 = \beta_7 - \sum_{j=1}^6 \beta_j \end{cases}$$

em ordem aos  $\beta_j$ , obtém-se

$$\begin{cases} \beta_j = \alpha_j - \bar{\alpha}, & \text{para } j=1, \dots, 6 \\ \beta_7 = \sum_{j=1}^7 \frac{\alpha_j}{7} = \bar{\alpha} \end{cases}$$

Como se pode constatar, a segunda especificação é uma simples reparametrização da primeira, pelo que os resultados obtidos serão sempre os mesmos. Note-se que em rigor a especificação (3.1) ou (3.2), para além de um efeito "puro dia da semana", contempla também um efeito "número de dias do mês", conclusão que resulta particularmente óbvia na reparametrização (3.2). Espasa (1982), na especificação do modelo de intervenção para a série mensal do IPI espanhol, opta pela representação da segunda alternativa dada a maior ortogonalidade que se consegue obter entre as variáveis  $T_{jt}$ .

Liu (1986) sugere ainda uma especificação mais complexa para captar o efeito "dias da semana" na qual, em vez de se assumir que o efeito associado a cada dia da semana é constante, admite que o mesmo cresce linearmente no tempo. Desta forma, ter-se-ia a seguinte especificação:

$$DS_t = \sum_{j=1}^7 \beta_j T_{jt} + \sum_{j=1}^7 \gamma_j (T_{jt} k_t) = \sum_{j=1}^6 (\beta_j + \gamma_j k_t) (Z_{jt} - Z_{7t}) + (\beta_7 + \gamma_7 k_t) \sum_{j=1}^7 Z_{jt} \quad (3.3)$$

em que  $k_t$  assume o valor 1 para as observações do primeiro ano, 2 para as do segundo ano, etc. Admite-se, deste modo, que o efeito "dia da semana" é constante dentro de cada ano, mas varia de ano para ano. Esta especificação pode ser considerada como uma tentativa de flexibilização da especificação (3.2). Todavia, a hipótese de cada coeficiente do efeito "dias da semana" crescer no tempo a uma taxa constante  $\gamma_j$  ( $j=1, \dots, 7$ ) poderá revelar-se tão restritiva quanto assumir que os parâmetros são constantes no tempo.

Da combinação dos modelos ARIMA univariados com os modelos de regressão, por via da introdução das variáveis "dias da semana", e admitindo um efeito "dias da semana" constante ao longo dos anos, obtém-se o seguinte modelo:

$$y_t = \sum_{j=1}^7 \alpha_j Z_{jt} + \frac{\Theta_q(B^*)\theta_q(B)}{\Delta^d \Delta_s^D \Phi_p(B^*)\phi_p(B)} \varepsilon_t$$

ou

$$y_t = \sum_{j=1}^7 \beta_j T_{jt} + \frac{\Theta_q(B^*)\theta_q(B)}{\Delta^d \Delta_s^D \Phi_p(B^*)\phi_p(B)} \varepsilon_t$$

consoante se utilize a primeira ou a segunda alternativa.

Hillmer (1982), através de uma aplicação empírica em que utiliza a mesma série temporal que serviu para a especificação de um modelo ARIMA univariado por Thompson e Tiao (1971), mostra que a inclusão das variáveis "dias da semana" potencia "ganhos" ao

nível da previsão. Na realidade, ao comparar o modelo obtido por Thompson e Tiao com o modelo em que são introduzidas as variáveis  $T_{jt}$  conclui que não só há uma melhoria em termos da qualidade do ajustamento mas, também, que com o modelo ARIMA "alargado" consegue-se uma redução substancial dos desvios padrão dos erros de previsão. Esta melhoria na capacidade preditiva do modelo que inclui as variáveis "dias da semana" deriva do facto de que, qualquer que seja o momento de previsão, estas variáveis são sempre conhecidas, dado que se conhece a composição do calendário.

No âmbito dos modelos estruturais, Harvey (1990) propõe a introdução de uma componente adicional ( $\tau_t$ ) para captar o efeito "dias da semana", obtendo-se o seguinte modelo estrutural "alargado":

$$y_t = \mu_t + \gamma_t + \tau_t + \varepsilon_t \quad (3.4)$$

em que  $\mu_t$ ,  $\gamma_t$ ,  $\varepsilon_t$  são, respectivamente, a componente tendência/ciclo, componente sazonal e componente irregular da série temporal  $y_t$ .

Para a componente  $\tau_t$  propõe a seguinte especificação:

$$\tau_t = \sum_{j=1}^6 \lambda_j (Z_{jt} - Z_{7t}) \quad (3.5)$$

em que  $Z_{jt}$  tem o mesmo significado que o definido anteriormente, isto é, o número de vezes que o dia  $j$  ( $j=1, \dots, 7$ ) ocorre no mês  $t$ ,  $\lambda_j$  é um parâmetro desconhecido associado ao dia  $j$  ( $j=1, \dots, 6$ ) e para  $j=7$  tem-se  $\lambda_7 = -\sum_{j=1}^6 \lambda_j$ .

Há dois aspectos que distinguem a especificação (3.5) das especificações (3.1) e (3.2).

O primeiro aspecto está relacionado com a exclusão, no modelo estrutural, da variável "número de dias do mês" (variável  $T_{jt}$  na especificação (3.2)). Esta omissão é válida no pressuposto de que os dados da série temporal foram previamente corrigidos no

sentido de se atenderem às diferenças de dimensão dos vários meses do ano<sup>[6]</sup>, ou seja, do efeito "número de dias do mês", ou então no caso de este se considerar incluído na componente sazonal mensal.

O segundo aspecto diz respeito à interpretação a dar aos coeficientes  $\lambda_j$  e  $\beta_j$  ( $j=1,\dots,6$ ). No modelo estrutural, a imposição da restrição de aditividade ( $\lambda_7 = -\sum_{j=1}^6 \lambda_j$ ), ou seja, da condição de que a componente "dias da semana" tem período 7, significa que

$$\tau_t = \sum_{j=1}^6 \lambda_j (Z_{jt} - Z_{7t}) = \sum_{j=1}^6 \lambda_j Z_{jt} + \left( -\sum_{j=1}^6 \lambda_j \right) Z_{7t} = \sum_{j=1}^6 \lambda_j Z_{jt} + \lambda_7 Z_{7t} = \sum_{j=1}^7 \lambda_j Z_{jt}$$

Deste modo, a especificação sugerida por Harvey permite que os parâmetros  $\lambda_j$  ( $j=1,\dots,6$ ) sejam interpretados como o efeito de cada um dos seis primeiros dias da semana sobre a variável  $y_t$ , sendo o efeito de Domingo dado por  $(-\sum_{j=1}^6 \lambda_j)$ . Comparando as especificações (3.2) e (3.5), e pondo de parte a questão relativa à variável "número de dias do mês", verifica-se que, apesar de não se ter imposto a restrição de aditividade em (3.2), os parâmetros  $\beta_j$  podem ser interpretados da mesma forma que os parâmetros  $\lambda_j$  ( $j=1,\dots,6$ ).

Uma condição que deve ficar garantida na modelação dos efeitos de "calendário", tal como no efeito sazonal, é a de que, atendendo ao seu carácter temporário, estes devem cancelar-se ao fim de um certo período (em geral, um ano), de forma a não afectarem o nível da tendência da série. No caso do efeito "dias da semana" esta condição está sempre verificada, pois  $(\sum_{t=1}^s Z_{jt} - \sum_{t=1}^s Z_{7t}) \approx 0$  para  $j=1,2,\dots,6$  e portanto:

$$\sum_{t=1}^s \tau_t = \sum_{t=1}^s \sum_{j=1}^6 \lambda_j (Z_{jt} - Z_{7t}) = \sum_{j=1}^6 \lambda_j (\sum_{t=1}^s Z_{jt} - \sum_{t=1}^s Z_{7t}) = 0, \quad s=4 \text{ ou } s=12$$

Repare-se que se o ano tivesse 364 dias, ao invés de ter 365 ou 366,  $\sum_{t=1}^s \tau_t = 0$ .

<sup>[6]</sup> Uma situação particular com interesse ocorre quando a variável  $y_t$  vem definida em termos de diferenças homólogas de logaritmos ou em taxas de variação homólogas. Neste caso, o coeficiente associado a  $T_{7t}$  acaba por captar os efeitos dos anos bissextos. Se se admitir que não existem efeitos adicionais decorrentes do ano ser ou não bissexto ou se se efectuar uma correcção prévia apenas nos dados dos meses de Fevereiro bissextos, então a exclusão da variável  $T_{7t}$  será também admissível.



Os procedimentos apresentados até ao momento permitem quantificar o efeito "dias da semana" para cada um dos dias da semana. Existem, contudo, situações concretas em que diferentes dias da semana produzem o mesmo efeito, ou seja,  $\lambda_i = \lambda_j$  ( $i \neq j$ ,  $i, j = 1, \dots, 7$ ). Para estes casos Harvey propõe uma especificação mais parcimoniosa para modelar o referido efeito, partindo da especificação obtida para captar os efeitos "diários"<sup>[6]</sup>.

Seja, então,  $\lambda_j^*$  o coeficiente associado ao conjunto dos dias da semana cujo efeito "dias da semana" é semelhante ( $j = 1, \dots, w$ ),  $k_j$  o número de dias da semana abrangidos pelo grupo  $j$  e  $n_{jt}^*$  o número total de dias do "tipo  $j$ " ( $j = 1, \dots, w$ ) existentes no mês  $t$ <sup>[7]</sup>. Pode-se especificar  $\tau_t$  da seguinte forma:

$$\tau_t = \sum_{j=1}^w \lambda_j^* n_{jt}^* = \sum_{j=1}^{w-1} \lambda_j^* n_{jt}^* + \lambda_w^* n_{wt}^* \quad (3.6)$$

Atendendo à periodicidade da componente "dias da semana", o efeito do  $w$ -ésimo "tipo" de dia é definido com base no pressuposto de que, para qualquer semana, o somatório dos efeitos diários deve anular-se. Desta forma ter-se-ia:

$$\sum_{j=1}^w \lambda_j^* k_j = 0 \Rightarrow \lambda_w^* = -\frac{1}{k_w} \sum_{j=1}^{w-1} \lambda_j^* k_j \quad (3.7)$$

Substituindo (3.7) em (3.6), obtém-se a seguinte especificação para a componente "dias da semana":

$$\tau_t = \sum_{j=1}^{w-1} \lambda_j^* n_{jt}^* + n_{wt}^* \left( -\frac{1}{k_w} \sum_{j=1}^{w-1} \lambda_j^* k_j \right) = \sum_{j=1}^{w-1} \lambda_j^* \left( n_{jt}^* - \frac{k_j n_{wt}^*}{k_w} \right) \quad (3.8)$$

---

<sup>[6]</sup> Harvey (1990), pp.43-44.

<sup>[7]</sup> Por exemplo, se o efeito "dias da semana" é idêntico de segunda a quinta-feira, mas distinto nos restantes dias da semana, então:  $w=4$ ,  $k_1=4$  e  $k_2=k_3=k_4=1$ .

Deste modo, a partir do conhecimento prévio de que existem determinados dias da semana cujo efeito é idêntico, ou seja, de que apenas existem  $w < 7$  efeitos distintos, pode-se conseguir uma redução do número de parâmetros a estimar da componente "dias da semana" de 6 para  $(w-1)$ .

As especificações apresentadas para modelar o efeito "dias da semana" são válidas na generalidade das situações, existindo, contudo, determinados casos em que não serão totalmente adequadas. Um exemplo desses casos ocorre quando se pretende modelar o referido efeito numa série temporal que mede o nível de actividade de uma unidade económica que apenas labora nos dias úteis da semana<sup>[8]</sup>. Outra questão relevante na modelação do efeito "dias da semana" diz respeito à quantificação do efeito "feriado" na série temporal. A existência de feriados no mês poderá distorcer o efeito "dias da semana", pelo que será necessário ter em conta o seu efeito. Harvey (1990), por exemplo, sugere que se considere o feriado como um oitavo dia da semana ou, então, conhecendo a natureza da série, como um dos dias da semana (por exemplo, um domingo). Espasa (1982) na modelação da série do IPI opta pela primeira sugestão, definindo uma variável adicional que, para um determinado mês, mede o número de feriados existentes nesse mês. Esta solução poderá também não ser totalmente aceitável dado que pressupõe que o efeito "feriado" é o mesmo qualquer que seja o dia da semana em que ele ocorra.

### III.1.2 - Detecção do Efeito "Dias da Semana"

A detecção do efeito "dias da semana" pode ser efectuada *a posteriori*, isto é, após se ter especificado e estimado o modelo no qual se incorpora à partida a componente "dias da semana" ou, então, antes da especificação de qualquer modelo para a série temporal em estudo, embora, na prática, a presença ou ausência deste efeito tende a ser habitualmente testada através da realização de inferência sobre determinados parâmetros dos modelos que já têm incorporada a componente "dias da semana".

---

<sup>[8]</sup> Numa situação destas o procedimento habitualmente sugerido consiste em corrigir previamente os dados da série, dividindo o total mensal pelo número de dias úteis do respectivo mês.

A título exemplificativo considere-se o seguinte modelo ARIMA "alargado" para uma série temporal que, em virtude da sua própria natureza, se supõe que esteja afectada pelo efeito "dias da semana":

$$y_t = \sum_{j=1}^6 \beta_j (Z_{jt} - Z_{7t}) + \beta_7 \sum_{j=1}^7 Z_{jt} + \frac{\Theta_Q(B^*)\theta_q(B)}{\Delta^d \Delta_s^D \Phi_P(B^*)\phi_p(B)} \varepsilon_t$$

Pierce (1971) mostrou que, assumindo a normalidade da série  $\{\varepsilon_t\}$ , os estimadores de  $(\beta, \Phi, \phi, \Theta, \theta)^{[9]}$  são consistentes e têm uma distribuição assintótica normal e que os estimadores de  $\beta$  são assintoticamente independentes dos estimadores de  $(\Phi, \phi, \Theta, \theta)$ . Além disso, também o estimador de  $\sigma_\varepsilon^2$  é consistente. Nesta situação, podem efectuar-se os seguintes testes de hipóteses para detectar a presença do efeito "dias da semana":

1<sup>o</sup> - testar a existência do efeito "puro dias da semana":

$$H_0: \beta_1 = \dots = \beta_6 = 0 \quad \text{versus} \quad H_a: \exists \beta_j \neq 0 \quad (j=1, \dots, 6)$$

A estatística de teste virá dada por:

$$\hat{\beta}_* A^{-1} \hat{\beta}_*$$

em que  $\hat{\beta}_* = (\hat{\beta}_1, \dots, \hat{\beta}_6)$  e  $A$  é a matriz de variâncias-covariâncias de  $\hat{\beta}_*$ . Para  $n$  grande  $\hat{\beta}_* \sim N(0, A)$ . Assim, sob a hipótese nula

$$\hat{\beta}_* A^{-1} \hat{\beta}_* \sim \chi^2(6)$$

2<sup>o</sup> - testar a existência do efeito "número de dias do mês"<sup>[10]</sup>:

$$H_0: \beta_7 = 0 \quad \text{versus} \quad H_a: \beta_7 \neq 0$$

A estatística de teste será o t-rácio do coeficiente estimado  $\hat{\beta}_7$ .

<sup>[9]</sup> Em que  $\beta = (\beta_1, \dots, \beta_7)$ ,  $\Phi = (\Phi_1, \dots, \Phi_p)$ ,  $\phi = (\phi_1, \dots, \phi_p)$ ,  $\Theta = (\Theta_1, \dots, \Theta_q)$  e  $\theta = (\theta_1, \dots, \theta_q)$ .

<sup>[10]</sup> Note-se que se a série  $y_t$  estiver expressa em variações homólogas, este teste equivale a testar a existência do efeito "anos bissextos".

Relativamente aos procedimentos que visam detectar o efeito "dias da semana" previamente à especificação de qualquer modelo é de referir os desenvolvidos por Cleveland e Devlin (1980)<sup>[11]</sup>. Estes autores propõem duas formas alternativas para detectar o referido efeito em séries temporais mensais: no domínio da frequência, sugerem a estimação da função de densidade espectral da série temporal, enquanto que no domínio tempo, aconselham o recurso a métodos gráficos. Para ambos os casos, sugerem que se eliminem previamente as componentes tendência/ciclo e sazonal, bem como que se corrija a componente irregular de quaisquer *outliers* detectados, de modo a tornar mais clara a detecção do efeito "dias da semana". Embora considerem os dois "grupos" de procedimentos como métodos complementares para a detecção do efeito, são peremptórios em afirmar que os métodos desenvolvidos no domínio da frequência são claramente mais poderosos que os desenvolvidos no domínio tempo.

No que se refere à função de densidade espectral, a detecção de picos em determinadas frequências do espectro da componente irregular da série — chamadas as "frequências de calendário" — é indicativa de que o efeito "dias da semana" está presente na respectiva série temporal. Espasa (1982), por exemplo, utiliza este procedimento para captar referido efeito na série mensal do IPI espanhol. Relativamente aos procedimentos no domínio tempo, Cleveland e Devlin utilizam determinados métodos gráficos desenvolvidos no âmbito do ajustamento sazonal para detectar o efeito "dias da semana". A detecção de determinados padrões de comportamento na componente irregular da série temporal, permite concluir que este efeito está presente.

---

<sup>[11]</sup> Neste trabalho, far-se-á apenas uma breve referência aos procedimentos apresentados por Cleveland e Devlin, atendendo a que saem fora dos objectivos do mesmo.

### III.2. - Efeito "Páscoa" — Detecção e Correção

O efeito "Páscoa" está presente numa série temporal quando, em virtude da ocorrência da "festa" Páscoa, se verifica uma alteração do nível da série nos meses em que ela ocorre. Consoante a natureza da respectiva série, essa alteração pode dar-se no sentido da expansão ou da contracção, podendo em determinadas situações verificarem-se ambos os casos. Como exemplos de séries temporais que pertencem ao primeiro grupo, ou seja, sobre as quais a Páscoa tenderá a ter um efeito expansivo, têm-se as vendas a retalho, o tráfego aéreo de passageiros, a circulação monetária, etc. No segundo grupo pode-se incluir, entre outras, a série do IPI (índice de produção industrial), uma vez que perante a ocorrência de um conjunto de feriados na semana da Páscoa é usual observar-se uma redução do nível da actividade industrial.

Em termos gerais, pode-se afirmar que a detecção do efeito "Páscoa" tenderá a ser mais imediata que a do efeito "dias da semana", pois uma análise, mesmo que superficial, do cronograma da série poderá ser o suficiente para informar quanto à presença ou ausência do efeito<sup>[12]</sup>. Atendendo a que a Páscoa só poderá ocorrer nos meses de Março e ou Abril — em termos trimestrais, no primeiro e ou segundo trimestres — uma alteração acentuada do nível da série nestes meses será indicativa de que o efeito "Páscoa" estará presente na série temporal.

As especificações sugeridas na literatura para modelar o efeito "Páscoa" aparecem normalmente enquadradas na modelação de uma componente mais genérica denominada "festas móveis". Na realidade, a necessidade em se corrigir o efeito "Páscoa" decorre do facto da Páscoa ser uma "festa móvel", ou seja, se esta "festa" ocorresse sempre no mesmo mês, o seu efeito teria um carácter sazonal e estaria, deste modo, contemplado na componente sazonal. Embora seja frequente associar o efeito "festas móveis" ao "efeito Páscoa" poderão existir, contudo, outras festas móveis cuja modelação poderá ser relevante, dependendo da série que se pretenda estudar. Atendendo a que, neste trabalho, o interesse recai sobre o efeito "Páscoa" ir-se-á modelar este efeito tendo presente que as especificações apresentadas são adequadas para modelar o efeito de qualquer outra festa móvel.

---

<sup>[12]</sup> Excepto se a série estiver "contaminada" por outras intervenções cujos efeitos compensem o efeito "Páscoa".

A abordagem tradicional para modelar o efeito "Páscoa" passa, primeiro que tudo, pela definição da duração e intensidade do efeito. Quanto ao primeiro aspecto, há que estabelecer o número de dias afectados pelo referido efeito e, tendo fixado este, há que determinar o coeficiente de ponderação associado a cada um desses dias. Como o que é relevante é determinar o peso relativo de cada um dos dias no efeito total, o somatório dos coeficientes de ponderação diários deve somar a unidade.

Desta forma, é usual propôr-se a seguinte especificação para o efeito "Páscoa" (PAS<sub>t</sub>):

$$PAS_t = \varphi H_t \quad (3.9)$$

em que a  $H_t$  é uma variável que assume valores não nulos apenas nos meses de Março e Abril e  $\varphi$  é um parâmetro a estimar.

Admitindo que o "período" da Páscoa compreende  $g$  dias, sendo  $g_1$  dias pertencentes ao mês de Março e  $(g-g_1)$  ao mês de Abril e que  $h_{jt}$  ( $j=1,\dots,g$ ) é o coeficiente de ponderação diário associado ao dia  $j$  do mês  $t$ , a variável  $H_t$  terá a seguinte estrutura:

$$H_t = \begin{cases} 0 & , \text{ para } t \neq \text{Março ou Abril} \\ \sum_{j=1}^{g_1} h_{jt} & , \text{ para } t = \text{Março} \\ \sum_{j=g_1+1}^g h_{jt} & , \text{ para } t = \text{Abril} \end{cases}$$

em que  $\sum_{j=1}^{g_1} h_{jt}$  e  $\sum_{j=g_1+1}^g h_{jt}$  medem, respectivamente, a proporção do efeito "Páscoa" respeitante aos meses de Março e Abril.

Na prática, a determinação de  $g$ ,  $g_1$  e dos  $h_{jt}$  ( $j=1,\dots,g$ ) efectua-se com base em informação que se dispõe a priori sobre o comportamento da série no "período" da Páscoa. Em séries temporais mensais ou trimestrais a indisponibilidade de informação diária leva a que, em muitos casos, os coeficientes de ponderação  $h_{jt}$  sejam fixados de uma forma meramente *ad hoc*. Por exemplo, Thury (1988) na especificação de modelos de intervenção

para as séries mensais do consumo total e das várias componentes do consumo admite um período de 10 dias afectado pelo efeito "Páscoa"<sup>[13]</sup>, atribuindo a cada um dos dias a proporção 0.10. A aceitação de um esquema de ponderações distribuídas uniformemente ao longo do período poderá ser discutível, pois raras serão as situações que isso se verificará. Espasa (1982) na modelação da série do IPI espanhol introduz também uma variável para captar o efeito "Páscoa", em que considera um período de oito dias, mas cujos coeficientes de ponderação diários são fixados por forma a reflectirem a redução do nível da actividade industrial em cada um dos dias da semana da Páscoa.

Na especificação (3.9) admite-se que o efeito "Páscoa" é constante ao longo dos anos. Contudo, em determinadas situações pode acontecer que o efeito aumente (diminua) segundo uma tendência linear crescente (decrescente). Liu (1986) considera que nesses casos será mais correcto admitir a seguinte especificação para modelar o efeito "Páscoa":

$$PAS_t = (\varphi_1 + \varphi_2 k_t) H_t \quad (3.10)$$

em que  $\varphi_1$  e  $\varphi_2$  são parâmetros a estimar e  $k_t$  e  $H_t$  têm o mesmo significado que o referido anteriormente.

Admitindo, para simplificar, que o efeito "Páscoa" é constante ao longo dos anos, isto é, que é utilizada a especificação (3.9), a série original corrigida deste efeito ( $y_t^*$ ) viria dada por

$$y_t^* = y_t - \varphi H_t$$

ou seja, por

$$y_t^* = \begin{cases} y_t & , \text{ para } t \neq \text{Março ou Abril} \\ y_t - \varphi \sum_{j=1}^{g_1} h_{jt} & , \text{ para } t = \text{Março} \\ y_t - \varphi \sum_{j=g_1+1}^g h_{jt} & , \text{ para } t = \text{Abril} \end{cases}$$

---

<sup>[13]</sup> Este período vai desde o Sábado anterior ao Sábado Santo até à segunda-feira imediatamente posterior ao Domingo de Páscoa.

Como se pode verificar, esta forma de corrigir a série original do efeito "Páscoa" conduz a uma alteração do próprio nível da série de modo que, para cada ano, tem-se

$$\sum_{t=1}^s y_t \neq \sum_{t=1}^s y_t.$$

Uma forma expedita para solucionar este "problema"<sup>[14]</sup>, ou seja, de corrigir o efeito "Páscoa" garantindo, ao mesmo tempo, que em cada ano se verifique a igualdade entre os níveis das séries original e corrigida pode ser encontrada em Escrivá e Sastre (1988). Como a Páscoa apenas ocorre nos meses de Março e ou Abril, estes autores propõem que o efeito "Páscoa" seja imputado na sua totalidade a um dos dois meses. Ao "deixar" de ser uma "festa móvel", dado que passa a "ocorrer" num determinado mês do ano, o efeito "Páscoa" será semelhante ao efeito das "festas" que ocorrem em data fixa (como por exemplo, o Natal) sendo, como tal, captado pela componente sazonal.

Para exemplificar a proposta de Escrivá e Sastre admita-se que Abril é o mês para o qual será transferido todo o efeito "Páscoa". Num determinado ano, a variável  $H_t$  assumirá uma das seguintes estruturas, consoante o momento de ocorrência do "período" da Páscoa nesse ano. Assim, se o efeito ocorrer todo em Abril, virá  $H_t = 0, \forall t, t=1, \dots, s$ , enquanto que se ocorrer todo em Março ter-se-á:

$$H_t = \begin{cases} 0, & \text{para } t \neq \text{Março ou Abril} \\ 1, & \text{para } t = \text{Março} \\ -1, & \text{para } t = \text{Abril} \end{cases}$$

Por sua vez, se o "período" da Páscoa ocorrer parte em Março e parte em Abril, a variável  $H_t$  vem dada por

$$H_t = \begin{cases} 0 & , \text{ para } t \neq \text{Março ou Abril} \\ h_M = \sum_{j=1}^{s_1} h_{jt}, & \text{ para } t = \text{Março} \\ - h_M & , \text{ para } t = \text{Abril} \end{cases}$$

onde os  $h_{jt}$  têm o mesmo significado referido anteriormente.

---

<sup>[14]</sup> Repare-se que esta questão pode ser entendida como um "problema" quando se está a trabalhar com a série em níveis. Em situações em que a série temporal utilizada está definida em variações homólogas esta questão torna-se irrelevante.



Harvey (1990), no âmbito dos modelos estruturais, propõe também uma especificação alternativa à especificação (3.9) em que, em termos anuais, fica garantida a igualdade dos níveis das séries  $y_t$  e  $y_t^*$ . Apresenta o seguinte modelo para captar o efeito "Páscoa":

$$PAS_t = \varphi H_t^* \quad (3.11)$$

$$\text{em que } H_t^* = \begin{cases} -\frac{1}{s} & , \text{ para } t \neq \text{Março ou Abril} \\ \sum_{j=1}^{g_t} h_{jt} - \frac{1}{s} & , \text{ para } t = \text{Março} \\ \sum_{j=g_1+1}^g h_{jt} - \frac{1}{s} & , \text{ para } t = \text{Abril} \end{cases}$$

Repare-se que parece haver alguma semelhança entre a especificação proposta por Harvey e a proposta por Escrivá e Sastre. Enquanto que estes dois últimos autores sugerem que o efeito seja totalmente transferido para um dos meses em que a Páscoa poderá ocorrer, Harvey propõe como que uma redistribuição do efeito "Páscoa" pelos vários meses do ano, por forma a que, em termos anuais, acabe por se anular. Atendendo a esta interpretação, de certa forma análoga, que se pode efectuar destas duas especificações apresentadas para corrigir o efeito "Páscoa", a proposta de Escrivá e Sastre parece ser mais adequada, uma vez que a correcção deste efeito, utilizando a especificação de Harvey, implica uma alteração do nível da série original em todos os meses do ano.

Refira-se ainda que a correcção do efeito "dias da semana" e do efeito "Páscoa" constitui uma exigência fundamental também ao nível dos métodos de dessazonalização de séries temporais. Na realidade, e como Hillmer (1982) o refere, os primeiros estudos realizados com o objectivo de captar este efeito visavam obter uma maior robustez dos resultados provenientes dos procedimentos de dessazonalização<sup>[16]</sup>. A não incorporação da componente "dias da semana" e da componente "Páscoa" no procedimento de decomposição de uma série temporal em que estes efeitos estarão presentes tenderá a originar um enviesamento das estimativas das restantes componentes, ou seja, da

---

<sup>[16]</sup> Entre eles menciona o de Young (1965) e do de Cleveland e Devlin (1980).

tendência/ciclo, componente sazonal e componente irregular, não se garantindo, desta forma, um ajustamento sazonal fiável.

Por esta razão, existem hoje determinados procedimentos de decomposição de séries temporais que, pelo menos nas suas versões mais recentes, incorporam a correcção dos "dias da semana" e do efeito "Páscoa". Como exemplo, tem-se o método de dessazonalização X-11 ARIMA (versão de 1988), desenvolvido no *Bureau of Statistics* do Canadá por Estella Dagum e, actualmente, considerado como um dos métodos mais aconselháveis para se proceder à dessazonalização de séries temporais. Este procedimento, cujo objectivo é a obtenção da série ajustada de sazonalidade, oferece a possibilidade de se corrigirem os efeitos de "calendário" (efeito "dias da semana" e "Páscoa") previamente ao cálculo dos factores sazonais.

A obtenção, pelo método X-11 ARIMA, dos coeficientes associados a cada um dias da semana resume-se, basicamente, na aplicação do método dos mínimos quadrados a um modelo de regressão (designado por *trading day regression*) cuja variável dependente corresponde a uma primeira estimativa da componente irregular "contaminada" pelo efeito "dias da semana". Esta primeira estimativa é obtida após se ter expurgado, da série original, as primeiras estimativas das componentes tendência/ciclo e sazonal, calculadas por aplicação de médias móveis. No caso do efeito "Páscoa", a correcção realiza-se com base numa especificação cujos argumentos são os valores de uma primeira estimativa da componente irregular referentes aos meses de Março e Abril e o número de dias que ocorrem entre o Domingo de Páscoa e o dia 22 de Março (data considerada o limite inferior no conjunto das possíveis datas de ocorrência do Domingo de Páscoa)<sup>[16]</sup>.

---

<sup>[16]</sup> Um maior desenvolvimento dos procedimentos subjacentes à estimação dos coeficientes dos efeitos "Páscoa" e "dias da semana" pode ser encontrado em Dagum (1988).

#### IV - APLICAÇÃO DA METODOLOGIA DA ANÁLISE DE INTERVENÇÃO À CORRECÇÃO DO INDICADOR DO CONSUMO CORRENTE

Os desenvolvimentos teóricos apresentados nos capítulos I, II e III, relativos à metodologia da análise de intervenção, detecção e correcção de observações anormais (*outliers*) e ajustamento dos efeitos "dias da semana" e "Páscoa" têm assumido um papel de grande relevo nos estudos empíricos realizados no âmbito das séries temporais. Na realidade, e como anteriormente foi referido, ao estudar uma determinada série temporal com base na metodologia ARIMA univariada é frequente concluir que este procedimento é, só por si, inadequado para modelar a série, em virtude de esta estar "contaminada" por um conjunto de fenómenos de natureza determinística não captados pelo modelo ARIMA univariado.

Neste capítulo aplica-se a análise de intervenção à série do indicador do consumo corrente<sup>[1]</sup>, tendo em vista ajustá-la de um conjunto de fenómenos de carácter exógeno e determinístico que ao não serem tidos em conta põem em causa a utilidade deste indicador para efeitos de análise de conjuntura. Far-se-á também a detecção e correcção de *outliers* por forma a que "choques" exógenos não captados pelas variáveis de intervenção não enviesem os resultados obtidos da estimação do modelo. Como foi referido no capítulo II, o programa que se utilizou para realizar a aplicação empírica — o SCA — tem incorporado um dos mais recentes procedimentos para detectar e corrigir observações anormais numa série temporal. Este procedimento, desenvolvido por Chen e Liu (1993), permite estimar conjuntamente e de uma forma automática os efeitos dos *outliers* e os parâmetros do modelo, pelo que neste trabalho o *output* relativo aos *outliers* refere-se ao obtido directamente pelo programa.

Para ajustar o indicador do consumo corrente foram consideradas duas hipóteses de correcção. Na primeira hipótese, o indicador actualmente publicado pelo Banco de Portugal foi corrigido de um conjunto de efeitos que, conhecendo a natureza da série e a forma como é construída, estariam a afectar o seu comportamento. Estes foram designados por efeito "dias da semana", efeito "Páscoa", efeito "rendimento" e efeito "novo hipermercado". Na segunda hipótese, e em virtude dos resultados não muito satisfatórios

---

<sup>[1]</sup> Esta série é objecto de publicação mensal pelo Banco de Portugal.

que se obtiveram na primeira, procedeu-se a uma reconstrução do indicador na qual se corrigiu *a priori* o efeito "novo hipermercado", tendo a metodologia da análise de intervenção sido aplicada ao "novo indicador" de modo a corrigi-lo dos restantes efeitos.

Na secção IV.1 começa-se por fazer uma breve referência à actual forma de cálculo do indicador do consumo corrente, seguindo-se a apresentação das variáveis de intervenção subjacentes à primeira hipótese e dos resultados empíricos obtidos. Por último, apresentar-se-á a segunda hipótese de correcção do indicador do consumo corrente e os respectivos resultados. Ao longo do texto far-se-á referência a gráficos e tabelas, parte dos quais se encontram em anexo.

#### IV.1 - O Indicador do Consumo Corrente

O indicador do consumo corrente é uma série temporal de periodicidade mensal habitualmente utilizada para efeitos de análise de conjuntura. Ao pretender reflectir o andamento do consumo privado num âmbito infra-anual é fundamental que o seu comportamento esteja expurgado de quaisquer fenómenos que o possam distorcer e tornar o indicador inadequado para o objectivo com que foi construído.

Este indicador é um índice calculado a partir das vendas nominais das principais cadeias de supermercados existentes em Portugal. Estas vendas, após serem deflacionadas por determinadas classes do índice de preços no consumidor (IPC), são convertidas em índice por aplicação da taxa de variação homóloga das vendas reais ao índice do mês homólogo (considerou-se Janeiro de 1982=100)<sup>[2]</sup>. Deste modo, o valor do índice no momento  $t$  ( $ICC_t$ ) virá dado por:

$$ICC_t = ICC_{t-12} \times \frac{VV_t}{VV_{t-12}}$$

em que  $VV$  é o volume de vendas total.

---

<sup>[2]</sup> Um maior desenvolvimento da metodologia de construção do indicador do consumo corrente pode ser encontrada em Santos (1986).

Um aspecto relevante no cálculo deste indicador é o de que a amostra mantém-se estável entre dois momentos de tempo homólogos consecutivos, ou seja, no momento  $t$  as vendas reais que entram no cálculo do indicador ( $VV_t$  e  $VV_{t,12}$ ) respeitam ao mesmo conjunto de unidades vendedoras. A necessidade em garantir esta estabilidade da amostra deriva do facto de se ter verificado ao longo do tempo a abertura de novas unidades e o encerramento de unidades que já estavam na amostra.

Quando da construção do indicador, decidiu-se que a inclusão de uma nova unidade no seu cálculo teria lugar ao fim de 18 meses, ou seja, entendeu-se que a taxa de crescimento das vendas da nova unidade só deveria afectar a taxa de crescimento do índice decorridos 18 meses da data da sua abertura. Isto porque, nos primeiros 6 meses de actividade, a nova unidade iria provavelmente registar volumes de vendas mensais muito elevados, resultantes quer de um certo efeito "novidade" que tenderia a ocorrer neste período, quer de eventuais campanhas promocionais que habitualmente são levadas a cabo na fase de implantação no mercado. Os restantes 12 meses referem-se aos meses necessários para se assegurar a estabilidade da amostra<sup>[3]</sup>.

---

<sup>[3]</sup> Para mais pormenores veja-se Dias (1987).

## IV.2 - Primeira Hipótese de Correção do Indicador do Consumo Corrente

A correção do indicador do consumo corrente tornou-se premente após se ter constatado um elevado grau de irregularidade na série das taxas de variação homólogas do índice. Como se pode verificar da figura 3, em anexo, é visível a elevada volatilidade subjacente a esta série de taxas que, num período de aproximadamente 8 anos (entre Janeiro de 1985 e Janeiro de 1993), assumiu valores entre -10.0 e 29.1 por cento.

Tendo como objectivo a eliminação dos efeitos de um conjunto de fenómenos que estariam eventualmente a contribuir para este andamento irregular das taxas de variação homólogas do índice, começou-se por se especificar um modelo para o indicador do consumo corrente no qual foram incorporadas variáveis de intervenção para captar os seguintes efeitos: efeito "dias da semana", efeito "Páscoa", efeito "rendimento" e efeito "novo hipermercado".

### IV.2.1 - Definição das variáveis de intervenção

Nesta secção efectua-se uma descrição de cada um dos efeitos, acima referidos, com base nos quais se definiram as respectivas variáveis de intervenção. É de referir que a detecção de alguns dos efeitos foi realizada a partir de informação diária, respeitante ao ano de 1991 e a Abril de 1992, sobre as vendas nominais de uma unidade que, em 1991, representavam cerca de 30 por cento das vendas totais. Atendendo a que esta unidade se enquadra no grupo das "grandes superfícies"<sup>[4]</sup>, é necessário uma certa precaução quando se pretender generalizar ao resto da amostra a interpretação dada a alguns dos efeitos.

---

<sup>[4]</sup> Segundo os universos definidos pela Nielsen, as "grandes superfícies" são unidades que possuem uma área de vendas superior ou igual a 2500 m<sup>2</sup>.

#### IV.2.1.1 - Efeito "Dias da Semana"

À partida é de esperar que a evolução das vendas nos supermercados seja diferenciada consoante o dia da semana em que o seu nível esteja a ser medido e, de facto, a análise da série diária, acima mencionada, permitiu concluir imediatamente que as quantidades vendidas nas sextas-feiras e nos sábados tendem a ser superiores às registadas nos restantes dias da semana. Por conseguinte, o volume de vendas num determinado mês irá ser função do número de vezes que cada dia da semana ocorre nesse mês.

Para captar o efeito "dias da semana" no indicador do consumo corrente propôs-se a especificação (3.1) vista na secção 1.1 do capítulo III<sup>[5]</sup>. Como se pôde verificar, esta especificação tem subjacente a hipótese de que cada dia da semana tem um efeito distinto. Contudo, em situações em que o número total de intervenções é muito elevado<sup>[6]</sup> será de todo vantajoso propôr à partida uma especificação mais parcimoniosa.

As figuras 4 a 7, em anexo, que visam representar semanas "normais" de vendas, isto é, semanas que se supõe não estarem afectadas por nenhum dos outros três efeitos, parecem confirmar uma certa homogeneidade das quantidades vendidas em quatro dias da semana (de segunda-feira a quinta-feira). Por outro lado, constata-se que as vendas mais altas da semana são registadas nos sábados, seguidas das dos domingos e das sextas-feiras.

Com base nesta informação propôs-se a seguinte especificação alternativa para modelar o efeito "dias da semana" na série do indicador do consumo corrente:

$$DS_t = \alpha \left( \sum_{j=1}^4 Z_{jt} \right) + \sum_{j=5}^7 \alpha_j Z_{jt}$$

---

<sup>[5]</sup> Optou-se por esta especificação, em detrimento da de Harvey, atendendo a que, para além do efeito "dias da semana puro", permite captar o efeito "dias do mês" ou, mais concretamente, o efeito "ano bissexto" uma vez que o modelo virá especificado em taxas de variação homólogas. Como adiante se verá, os resultados obtidos confirmaram a relevância deste efeito.

<sup>[6]</sup> Como mais à frente se verá, será este o caso desta primeira hipótese de correcção.

expressão que se obtém de (3.1) admitindo que  $\alpha_1=\alpha_2=\alpha_3=\alpha_4=\alpha$ . Nesta especificação, a variável  $\sum_{j=1}^4 Z_{jt}$  mede o número total de segundas, terças, quartas e quintas-feiras que ocorrem no mês  $t$ .

#### IV.2.1.2 - Efeito "Páscoa"

Da análise da figura 8 referente às vendas diárias registadas no mês de Abril de 1992, mês em que a Páscoa ocorreu, detectaram-se dois tipos de efeitos associados ao efeito "Páscoa". Um efeito "expansivo" nas vendas, durante um período de sete dias, compreendido entre o sábado anterior ao Sábado Santo e a Sexta-Feira Santa, e que corresponde ao acréscimo esperado nas vendas resultante do facto de, durante esta quadra festiva, os consumidores aumentarem as suas compras para níveis acima dos normais, e um efeito "depressivo" nas vendas, no período que vai do Sábado Santo à quarta-feira da semana seguinte à da Páscoa, entendendo-se este como um período cujas vendas estão abaixo do nível "normal". Existem dois aspectos que podem conjuntamente explicar a contracção nas vendas observada neste período. Relativamente à quebra verificada no fim de semana da Páscoa, esta ficará a dever-se não só ao facto de nesta altura muitas pessoas saírem das grandes cidades (Lisboa e Porto) como, também, de estes dias não serem encarados como adequados à realização de compras por parte dos consumidores<sup>[7]</sup>. No que se refere à contracção das vendas verificada na segunda, terça e quarta-feira imediatamente posteriores ao Domingo de Páscoa, esta poderá resultar de um efeito "antecipação" verificado na semana imediatamente anterior.

Tendo-se detectado estes dois efeitos associados ao "período" da Páscoa, definiram-se duas variáveis de intervenção —  $PASP_t$  e  $PASN_t$  — associadas, respectivamente, ao efeito "expansivo" e ao efeito "depressivo". Estas duas variáveis foram construídas atribuindo um peso a cada um dos dias dos períodos "expansivo" e "depressivo". Estes pesos, no caso do efeito "expansivo" nas vendas<sup>[8]</sup>, foram calculados de acordo com a expressão:

---

<sup>[7]</sup> Note-se que a maior parte dos supermercados e hipermercados que pertencem à amostra estão localizados nestas duas grandes cidades.

<sup>[8]</sup> O mesmo tipo de cálculos foi efectuado para determinar os pesos diários do efeito "depressivo".



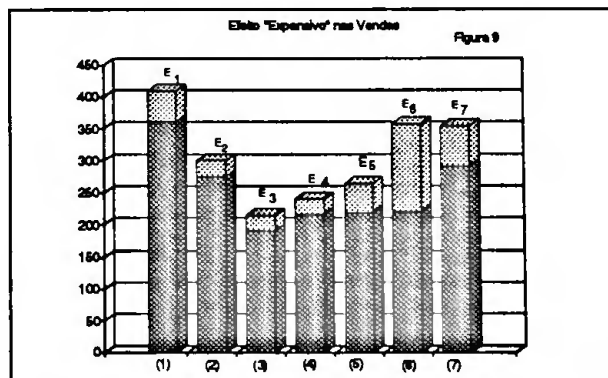
$$w_j^E = \frac{E_j}{\sum_{j=1}^7 E_j}$$

onde  $w_j^E$  é o coeficiente de ponderação associado ao dia  $j$  e  $E_j$  representa a diferença entre as vendas nominais registadas no dia  $j$  do período "expansivo" e as registadas no mesmo dia  $j$  de uma semana considerada "normal" ( $j=1, \dots, 7$ )<sup>[9]</sup>

Atendendo a que a Páscoa só poderá ocorrer em Março e ou em Abril, a estrutura da variável  $PASP_t$ , para um determinado ano em que  $g_1$  dias do efeito "expansivo" ocorrem em Março, virá dada por:

$$\begin{cases} \sum_{j=1}^{g_1} w_j^E, & \text{em Março} \\ \sum_{j=g_1}^7 w_j^E, & \text{em Abril} \\ 0, & \text{nos restantes meses} \end{cases}$$

Nas figuras 9 e 10 comparam-se, respectivamente, as vendas registadas nos períodos "expansivo" e "depressivo" da Páscoa com as vendas registadas em dias cujo nível de vendas é considerado normal. Para cada um dos casos ter-se-á a seguinte relação:  $E_j > 0$  ( $j=1, \dots, 7$ ) e  $D_j < 0$  ( $j=1, \dots, 5$ ). Nas tabelas 1.1 e 1.2 indicam-se os coeficientes de ponderação diários associados a cada um dos efeitos.



<sup>[9]</sup> Para simplificar a notação atribuiu-se um número a cada dia do período "expansivo". Assim, por exemplo,  $j=1$  corresponderá ao primeiro dia do "período", ou seja, ao sábado anterior ao Sábado Santo.

Tabela 1.1 - Efeito "expansivo"	
Dia	$w_j^E$
Sábado (1)	0.13
Domingo (2)	0.07
Segunda-feira (3)	0.06
Terça-feira (4)	0.07
Quarta-feira (5)	0.13
Quinta-feira Santa (6)	0.37
Sexta-feira Santa (7)	0.17
Total	1

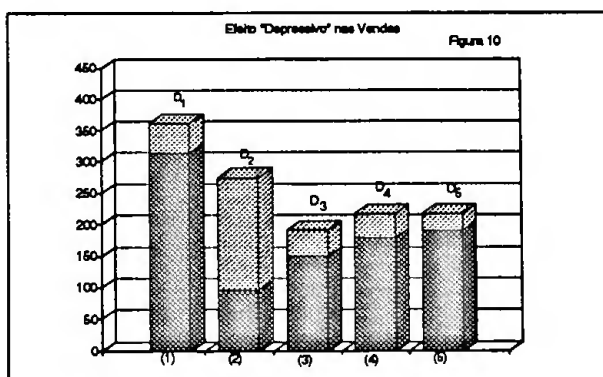


Tabela 1.2 - Efeito "depressivo"	
Dia	$w_j^D$
Sábado Santo (1)	0.14
Domingo de Páscoa (2)	0.53
Segunda-feira (3)	0.12
Terça-feira (4)	0.12
Quarta-feira (5)	0.09
Total	1

#### IV.2.1.3 - Efeito "Rendimento"

Subjacente a este efeito está a ideia de que durante o mês existe um período no qual a maior parte das pessoas terão restrições de liquidez uma vez que são remuneradas mensalmente. Atendendo a que parece existir uma certa preferência pelas sextas-feiras, sábados e domingos para realizarem as suas compras, segundo o que foi detectado nas figuras 4 a 7 do efeito "dias da semana", é de se esperar um acréscimo acima do normal das suas aquisições no "fim de semana" imediatamente posterior ao recebimento do seu vencimento<sup>[10]</sup>.

A detecção deste efeito ocorreu após se terem analisado as vendas nominais diárias registadas em 1991, tomando como período de referência o mês. Uma análise mais atenta da mesma informação permitiu ainda fixar o dia 28 de cada mês como a data a partir da qual as pessoas eventualmente teriam disponível o seu vencimento para realizarem despesas. Admitindo que em cada mês existe, em média, uma sexta-feira de "rendimento", um sábado de "rendimento" e um domingo de "rendimento"<sup>[11]</sup>, consideraram-se como anómalas as situações em que estes "dias" ocorrem duas ou zero vezes no mês.

Uma análise das figuras 11 a 13, em anexo, permite constatar que o mês de Janeiro não contém nenhum "fim de semana de rendimento", o que significa que logo no início do mês de Fevereiro ocorre um destes "fins de semana". Por outro lado, verifica-se que o mês de Agosto contém "quase" dois "fins de semana de rendimento".

É de referir que na definição das variáveis de intervenção associadas a este efeito houve o cuidado em verificar se depois do dia 28 de cada mês existia ou não um feriado. Relativamente aos feriados que ocorrem antes do dia 28, optou-se por ignorá-los no pressuposto de que apenas provocam uma antecipação das vendas, ou seja, de que os consumidores que habitualmente efectuam as suas compras na sexta-feira ou no sábado antecipam as suas aquisições para o dia anterior ao feriado ou para o dia do feriado<sup>[12]</sup>.

---

<sup>[10]</sup> Por uma questão de conveniência, designa-se "fim de semana" ao período que compreende a sexta-feira, o sábado e o domingo.

<sup>[11]</sup> Por sexta-feira de "rendimento", sábado de "rendimento" e domingo de "rendimento" quer-se designar, respectivamente, uma sexta-feira, um sábado e um domingo afectados pelo efeito "rendimento". Por analogia, designar-se-á um "fim de semana" afectado por esse mesmo efeito de "fim de semana de rendimento"

<sup>[12]</sup> Repare-se que, por outro lado, não será assim tão grave ignorar estes feriados uma vez que a maior parte deles ocorre em data fixa. Como tal, em termos homólogos, os seus efeitos anular-se-ão. Refira-se também que foram também ignorados os feriados que, mesmo tendo ocorrido após o dia 28, estivessem incluídos no "período

Deste modo, estes dois dias podem ser tomados, respectivamente, como uma sexta-feira e um sábado.

Quando o feriado ocorre após o dia 28 do mês esta neutralidade do efeito "feriado" poderá não estar presente. Na realidade, ainda que um determinado mês não possua uma sexta-feira de "rendimento" ou um sábado de "rendimento" mas contenha um feriado no final do mês pode, mesmo assim, possuir um "fim de semana de rendimento". Como exemplo de uma destas situações considerem-se as figuras 14 e 15, em anexo, onde estão representadas as vendas nominais diárias referentes aos meses de Maio e Junho de 1991. Tendo presente o conceito de "fim de semana de rendimento", a inexistência de nenhum dos dias do efeito "rendimento" no mês de Maio levaria a admitir a ausência deste efeito neste mês. No entanto, a ocorrência de um feriado no dia 30 permite que se contabilize um "fim de semana de rendimento" em Maio. De facto, note-se que o nível de vendas do primeiro sábado do mês de Junho não é comparável ao do último sábado do mesmo mês, que é um sábado de "rendimento". Em anexo apresentam-se as possíveis formas de captar o efeito "rendimento" num mês com 31 dias que poderá conter ou não um feriado após o dia 28.

Refira-se ainda que o feriado do Carnaval não foi considerado no grupo dos feriados que podem ocorrer após o dia 28 do mês, em virtude de se ter detectado um decréscimo de vendas na terça-feira de Carnaval que, de certa forma, acabou por compensar o acréscimo de vendas ocorrido no dia anterior (segunda-feira).

Com base no que acima ficou dito, definiram-se três variáveis de intervenção para captar o efeito "rendimento":  $SEXR_t$ ,  $SABR_t$  e  $DOMR_t$ , onde, por exemplo,  $SABR_t$  mede o número de sábados de "rendimento" e ou de feriados que ocorrem após o dia 28 do mês  $t$ , menos 1<sup>[13]</sup>.

---

expansivo" ou "depressivo" das vendas do efeito "Páscoa".

<sup>[13]</sup> Pela mesma lógica,  $SEXR_t$  mede o número de sextas-feiras de "rendimento" e ou de dias que antecedem os feriados que ocorrem depois do dia 28 do mês  $t$ , menos 1, enquanto que  $DOMR_t$  mede o número de domingos de "rendimento" existentes nesse mesmo mês, menos 1.

#### IV.2.1.4 - Efeito "Novo Hipermercado"

O efeito "novo hipermercado" visa captar o impacto da abertura das grandes superfícies sobre o índice do consumo corrente. Na realidade, o surgimento das grandes superfícies, verificado após 1986, teve fortes implicações no indicador a partir dessa data. Dadas as acentuadas diferenças estruturais existentes entre as unidades inicialmente incluídas na amostra (supermercados) e estas novas unidades (hipermercados) tornou-se necessário modelar os efeitos destas últimas no indicador.

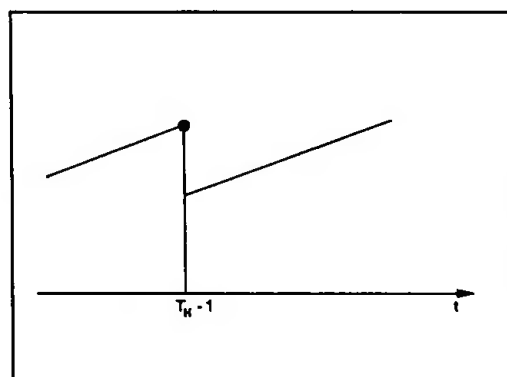
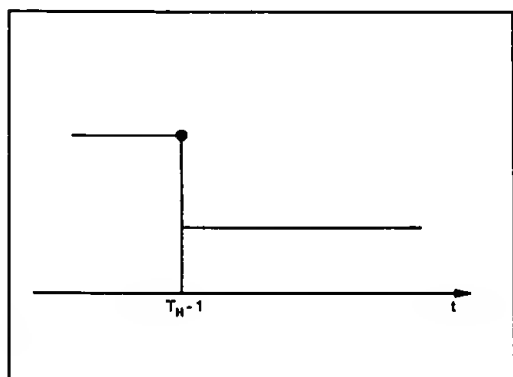
Como foi referido anteriormente, quando uma nova unidade, supermercado ou hipermercado, surge no mercado, as suas vendas só serão incluídas na amostra 6 meses após a sua abertura. Isto significa que a taxa de crescimento das vendas desta nova unidade só terá um impacto directo sobre a taxa de variação homóloga do índice do consumo corrente decorridos 18 meses após a sua entrada em funcionamento. Embora com esta regra se tenha procurado obter uma certa estabilidade da amostra e, ao mesmo tempo, evitar uma elevada volatilidade no andamento do indicador, o que é certo é que ela acabou se revelar bastante limitativa. Isto porque a abertura da nova unidade acaba por exercer uma influência sobre o índice através de dois canais, que poderão ser designados por efeito "desvio de comércio" e por efeito "*overshooting*".

No que se refere ao desvio de comércio, é habitual verificar-se, após a abertura de uma nova unidade num local onde já existem outras unidades em funcionamento, uma transferência de clientes, que anteriormente efectuavam as suas compras nessas unidades já existentes, para a nova unidade, pela simples razão de, por exemplo, se localizar a uma menor distância. Por outro lado, essa unidade, sobretudo se for um hipermercado, tende a possuir, pelo menos nos primeiros meses após a sua abertura, um poder de atracção superior ao das outras unidades já existentes. Esta atracção pelos produtos da nova unidade, que é usual verificar-se na prática, poderá ser consequência quer de um certo efeito "novidade" exercido pela nova loja, quer de eventuais promoções que são habitualmente levadas a cabo nesta fase de implantação no mercado. Deste modo, espera-se que a abertura da nova unidade origine algum desvio de comércio das outras unidades já existentes para esta nova unidade, sobretudo das que estão localizadas nas proximidades.

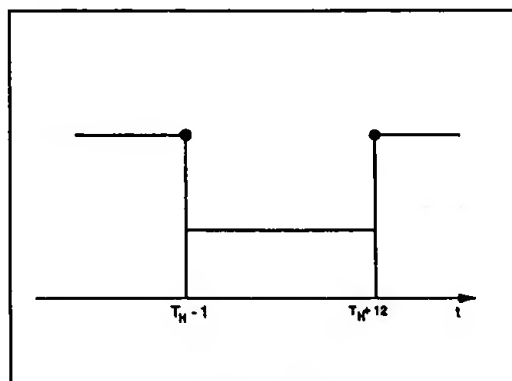
Consequentemente, se se atender à forma de cálculo do indicador do consumo corrente e ao facto de que estas outras unidades, afectadas pelo fenómeno do desvio do comércio, já estarem incluídas na amostra, a taxa de crescimento homóloga do indicador estará provavelmente a subestimar a "verdadeira" taxa de crescimento do consumo privado durante pelo menos 12 meses.

Para modelar o efeito "desvio de comércio" começou-se por considerar a hipótese de que o desvio de comércio se efectua de uma vez por todas, ou seja, de que assume um carácter permanente logo à partida. Em termos gráficos, podem ter-se os seguintes esquemas para representar o impacte da abertura do novo hipermercado H, no momento  $t=T_H$ , sobre:

\* o nível de vendas total — considerando os casos mais simples em que as vendas das unidades que estão na amostra estão a crescer a uma taxa nula ou constante ( $> 0$ ) tem-se, respectivamente:



\* as taxas de variação homólogas do índice — em ambas as situações acima representadas o efeito sobre as taxas é o seguinte:



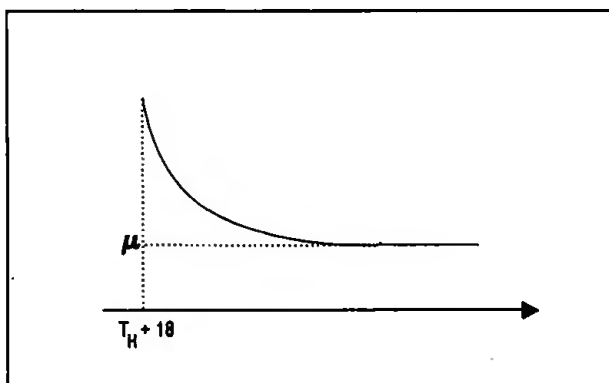
Neste caso, a variável de intervenção associada à abertura do novo hipermercado terá a seguinte estrutura:

$$DA_{H_t}^{[14]} = \begin{cases} 1, & \text{para } T_H \leq t \leq T_H+11 \\ 0, & \text{para os outros casos} \end{cases}$$

em que  $T_H$  é a data de abertura do hipermercado H.

No período amostral considerado para efeitos de estimação — 1984:7 a 1993:1 — registou-se a abertura de nove hipermercados, pelo que H assume valores entre 1 e 9. Atendendo a que os hipermercados foram ordenados por ordem crescente da sua data de abertura, H=1 designará o primeiro hipermercado que abriu, H=2 o segundo, e assim por diante.

Considere-se agora o efeito *overshooting*. Como foi referido anteriormente, a taxa de variação homóloga das vendas da nova unidade só afecta a taxa de crescimento do indicador 18 meses após sua abertura. Contudo, se as vendas do novo hipermercado, no momento  $t=T_H+18$ , estiverem a crescer a uma taxa muito mais elevada que as vendas das restantes unidades, a inclusão desta nova unidade no cálculo do indicador irá ter um efeito *overshooting* sobre a taxa de variação homóloga do índice, ou seja, fará com que esta "salte" para valores mais elevados que os anteriormente registados<sup>[16]</sup>. Tendo analisado as vendas de alguns hipermercados desde a sua abertura, detectou-se o seguinte padrão de evolução nas suas taxas de crescimento homólogas:



<sup>[14]</sup> Repare-se que esta variável é uma variável de intervenção do tipo *extending pulse*.

<sup>[16]</sup> A magnitude do impacto irá também depender do peso do volume de vendas do novo hipermercado no volume de vendas total.

Desta forma, verifica-se que quando o hipermercado H entra no cálculo do índice — no momento  $t=T_H+18$  — a taxa de crescimento das suas vendas tende a ser anormalmente elevada. Contudo, ela vai-se amortecendo exponencialmente ao longo do tempo a uma taxa  $\delta_H$ , convergindo, eventualmente, para a taxa de crescimento média ( $\mu$ ). Este comportamento levou a que se propusesse a seguinte estrutura para a variável de intervenção que capta o efeito "overshooting":

$$DEN_{H_t} = \begin{cases} w_{0H} \delta_H^{[t - (T_H + 18)]}, & \text{para } t \geq (T_H + 18) \\ 0 & , \text{ para os outros casos} \end{cases}$$

Repare-se que esta representação é equivalente a admitir-se uma variável "impulso" para a variável de intervenção  $DEN_{H_t}$  e a seguinte função transferência para medir o impacto sobre a variável *output*:

$$v_H(B) = \frac{w_{0H}}{1 - \delta_H B} = w_{0H}(1 + \delta_H B + \delta_H^2 B^2 + \delta_H^3 B^3 + \dots)$$

onde  $DEN_{H_t}$  é agora igual 1, para  $t=T_H + 18$ , e 0 nos restantes casos. Na expressão anterior,  $w_{0H}$  mede o impacto imediato sobre a série *output* e  $\delta_H$  a taxa à qual se amortece o efeito "overshooting" associado ao hipermercado H. Se  $\delta_H$  estiver próxima de zero significa que a taxa de crescimento do volume de vendas do novo hipermercado tem apenas um efeito temporário, de muito curta duração, na taxa de crescimento do índice. Se  $\delta_H$  estiver próxima de 1, nesse caso o efeito tende a ser permanente. Atendendo a que, até Janeiro de 1993, apenas as vendas dos cinco primeiros hipermercados que abriram tinham entrado no cálculo do índice, tem-se  $H=1, \dots, 5$ .

Esta especificação tem a vantagem de ser bastante geral, na medida em que contempla um conjunto de casos possíveis: admite um impacto nulo ( $w_{0H}=0$ ), um impacto que se esgota no momento em que hipermercado entra no cálculo do índice ( $\delta_H=0$ ) ou que se esgota ao longo do tempo ( $0 < \delta_H < 1$ ). Por último, contempla ainda o caso em que ocorre uma alteração do próprio nível da série das taxas com a entrada no indicador do novo hipermercado ( $\delta_H=1$ ).



Em resumo, na primeira hipótese de correcção do indicador do consumo corrente foram consideradas as seguintes variáveis de intervenção:

- efeito "dias da semana":  $\sum_{j=1}^4 Z_{jt}$ ,  $Z_{5t}$ ,  $Z_{6t}$  e  $Z_{7t}$ .
- efeito "Páscoa":  $PASP_t$  e  $PASN_t$ .
- efeito "rendimento":  $SEXR_t$ ,  $SABR_t$  e  $DOMR_t$ .
- efeito "novo hipermercado":  $DA_{H_t}$  ( $H=1,...,9$ ) e  $DEN_{H_t}$  ( $H=1,...,5$ ).

#### IV.2.2 - Resultados Empíricos

Previamente à apresentação dos resultados empíricos é importante realçar que enquanto que as variáveis associadas ao efeito "novo hipermercado" foram definidas para medir o impacto sobre as taxas de variação homólogas do índice do consumo corrente, as restantes variáveis foram definidas para captar o efeito sobre os níveis de vendas, ou seja, directamente sobre o índice. Com o intuito de tornar compatíveis estas duas formas, aparentemente distintas, de captar os efeitos definiram-se as variáveis de intervenção associadas aos efeitos "dias da semana", "Páscoa" e "rendimento" em termos de diferenças homólogas. Consequentemente, no modelo de intervenção a variável *output* virá dada por:

$$\left( \frac{ICC_t}{ICC_{t-12}} - 1 \right) = \Delta_{12} LICC_t$$

em que  $LICC_t$  é o logaritmo do índice do consumo corrente ( $ICC_t$ ).

Tendo presente esta transformação de variáveis, especificou-se o seguinte modelo de análise intervenção para modelar o comportamento das variações homólogas do indicador do consumo corrente:

$$\begin{aligned}
\Delta_{12}LICC_t = & \alpha_0 + \alpha \Delta_{12}\left(\sum_{j=1}^4 Z_{jt}\right) + \alpha_5 \Delta_{12}Z_{5t} + \alpha_6 \Delta_{12}Z_{6t} + \alpha_7 \Delta_{12}Z_{7t} + \\
& + \varphi_1 \Delta_{12}PASP_t + \varphi_2 \Delta_{12}PASN_t + \lambda_1 \Delta_{12}SEXR_t + \lambda_2 \Delta_{12}SABR_t + \\
& + \lambda_3 \Delta_{12}DOMR_t + \sum_{H=1}^9 \eta_H DA_H + \sum_{H=1}^5 \frac{\omega_{0H}}{1 - \delta_H B} DEN_H + \zeta_t
\end{aligned} \quad (4.1)$$

em que  $\alpha_0$  é o termo independente do modelo e  $\zeta_t$  a componente residual.

Atendendo à forma como se interpretou cada uma das variáveis de intervenção, espera-se que os parâmetros da componente sistemática do modelo tenham os seguintes sinais e que verifiquem as seguintes restrições:

Efeito	Parâmetros
"Dias da semana"	$\alpha, \alpha_i > 0$ ( $i=5,6,7$ ): $\alpha \leq (\alpha_5 \wedge \alpha_7) < \alpha_6$
"Páscoa"	$\varphi_1 > 0 \wedge \varphi_2 < 0$
"Rendimento"	$\lambda_i > 0$ ( $i=1,2,3$ ): $(\lambda_1 \wedge \lambda_3) < \lambda_2$
"Novo hipermercado"	$\eta_H < 0$ ( $H=1,\dots,9$ ) $\omega_{0H} \geq 0, 0 \leq \delta_H \leq 1$ ( $H=1,\dots,5$ )

Embora o índice do consumo corrente esteja disponível mensalmente desde Janeiro de 1982, optou-se por estimar o modelo (4.1) utilizando um período amostral mais curto, em virtude de se ter observado um comportamento sistematicamente negativo das taxas de variação homólogas do índice no período que antecede o segundo semestre de 1985<sup>[16]</sup>. Ainda que este comportamento possa ser explicado enquadrado na conjuntura recessiva que a economia portuguesa atravessou no período de 83-84, preferiu-se mesmo assim ignorá-lo, além de que, como mais adiante se verá, foi também este o período utilizado na segunda hipótese de correcção do indicador. Assim, o período de estimação usado inclui as observações desde 1984:7 a 1993:1 (103 observações), embora, após a aplicação de uma diferença de ordem 12, tenham ficado disponíveis 91 observações.

<sup>[16]</sup> Veja-se a figura 3.

No intuito de identificar um modelo inicial para a componente residual começou-se por estimar o modelo (4.1) incorporando a detecção e o ajustamento de observações atípicas. Para este efeito considerou-se um valor crítico de 2.7, pelas razões apontadas no capítulo II. Tendo presente o comportamento das taxas de variação homólogas do indicador, parece ser este o caso da série em estudo. Nas figuras 16 e 17, em anexo, estão representadas, respectivamente, a função de autocorrelação — FAC — e a função de autocorrelação parcial — FACP — estimadas da série residual ajustada de *outliers*<sup>[17]</sup>. Após se terem analisado estas funções propôs-se o seguinte modelo inicial para a série residual:

$$\zeta_t = (1 - \theta_6 B^6) \varepsilon_t \quad (4.2)$$

Tendo substituído (4.2) em (4.1) e estimado conjuntamente os parâmetros de ambas as componentes do modelo o coeficiente estimado de  $\theta_6$  veio não significativo pelo que se optou por não propôr nenhum modelo inicial para a série residual. Note-se contudo que, uma vez que os coeficientes de autocorrelação da primeira estimativa da componente residual não se afastam muito dos coeficientes de autocorrelação de uma série ruído branco, a não especificação de um modelo inicial para  $\zeta_t$  não terá, em princípio, repercussões graves ao nível da eficiência dos estimadores dos coeficientes das variáveis de intervenção. Refira-se, no entanto, que este comportamento poderá ficar a dever-se ao elevado número de parâmetros a estimar da componente sistemática (29 parâmetros!).

Nas tabelas 1.3.A a 1.3.C, que se apresentam nas páginas seguintes, indicam-se os resultados empíricos relativos ao modelo (4.1), os quais foram obtidos mediante a aplicação do procedimento sequencial-iterativo descrito na secção referente à estimação e avaliação do diagnóstico dos modelos de intervenção (secção I.2.1.2)<sup>[18]</sup>. Nesta fase eliminaram-se do modelo os parâmetros cujas estimativas vieram estatisticamente não significativas ou que apresentaram sinal contrário ao esperado. Em virtude da eliminação de algumas variáveis de intervenção houve necessidade de melhorar a especificação para a componente residual, pelo que, na especificação "final" se reteve um modelo MA(3)<sup>[19]</sup>.

<sup>[17]</sup> Esta série encontra-se representada na figura 18.

<sup>[18]</sup> A série residual ajustada de *outliers* e as funções de autocorrelação e autocorrelação parcial desta mesma série estão representadas, respectivamente, nas figuras 19, 20 e 21. Em anexo está também indicada a matriz de correlações dos parâmetros estimados — tabela 1.3.D.

<sup>[19]</sup> A estimativa do parâmetro média móvel consta também da tabela 1.3.A.

A partir destes resultados empíricos pode-se concluir que, relativamente à componente sistemática do modelo, apenas se confirmaram algumas das expectativas existentes quanto aos efeitos das intervenções sobre a série do indicador do consumo corrente.

Especificação "Final" do Modelo (4.1)				
Tabela 1.3.A	Parâmetro	Variável	Estimativa	t-rácio
1	$\alpha_0$		0.0308	14.94
2	$\alpha$	$\Delta_{12} \sum_{j=1}^4 Z_{jt}$	0.0158	1.77
3	$\alpha_6$	$\Delta_{12} Z_{6t}$	0.0133	1.36
4	$\alpha_6$	$\Delta_{12} Z_{6t}$	0.0519	5.80
5	$\alpha_7$	$\Delta_{12} Z_{7t}$	0.0046	0.51
6	$\varphi_1$	$\Delta_{12} \text{PASP}_t$	0.0459	5.71
7	$\varphi_2$	$\Delta_{12} \text{PASN}_t$	-0.0302	- 3.16
8	$\omega_1$	$\text{DE}_{1t}$	0.1803	30.85
9	$\delta_1$	$\text{DE}_{1t}$	0.9632	296.64
10	$\eta_2$	$\text{DA}_{2t}$	-0.0425	- 6.58
11	$\omega_2$	$\text{DE}_{2t}$	0.0974	9.84
12	$\delta_2$	$\text{DE}_{2t}$	0.9066	63.41
13	$\omega_4$	$\text{DE}_{4t}$	0.0652	4.36
14	$\eta_6$	$\text{DA}_{6t}$	-0.0178	- 3.66
15	$\eta_7$	$\text{DA}_{7t}$	-0.0117	- 2.17
16	$\eta_9$	$\text{DA}_{9t}$	-0.0247	- 2.82
17	$\theta_3$		0.5319	5.08
S.E. (sem ajustamento de outliers) = 0.042				
S.E. (com ajustamento de outliers) = 0.016				
Modelo da componente residual: $\zeta_t = (1 - \theta_3 B^3) \varepsilon_t$				

Tabela 1.3.B - Detecção e Ajustamento de *Outliers*

Momento de Ocorrência	Estimativa	t-rácio	Tipo de <i>outlier</i>
24	-0.051	- 4.98	AO
26	0.040	14.49	LS
43	0.071	4.34	IO
57	0.061	3.75	IO
75	-0.018	- 6.08	LS
79	0.087	5.19	IO
85	0.174	10.09	IO
87	0.082	7.87	AO
105	-0.057	- 3.92	AO

Tabela 1.3.C - Coeficientes de Autocorrelação da Série Residual

$\tau$	1	2	3	4	5	6	7	8	9	10	11	12
$\rho(\tau)$	-.07	-.12	-.14	-.19	-.03	.05	-.12	-.16	-.14	.09	.04	-.01
$\tilde{Q}(\tau)$	.4	1.9	3.7	7.4	7.4	7.7	9.1	11.7	13.7	14.5	14.7	14.8
$\tau$	13	14	15	16	17	18	19	0	21	22	23	24
$\rho(\tau)$	-0.2	.11	-.01	.04	.08	.01	.01	.01	-.16	-.02	-.13	.05
$\tilde{Q}(\tau)$	14.8	16.2	16.2	16.3	17.1	17.1	17.2	17.2	20.1	20.2	22.4	22.7
$\tau$	25	26	27	28	29	30	31	32	33	34	35	36
$\rho(\tau)$	.05	-.08	0.1	.03	.02	.10	-.01	.02	-.02	.06	.06	-.05
$\tilde{Q}(\tau)$	23.0	23.8	23.9	24.0	24.0	25.3	25.4	25.4	25.9	26.5	27.0	27.5

em que  $\rho(\tau)$  é a estimativa do coeficiente de autocorrelação de ordem  $\tau$  ( $\tau=1,\dots,36$ ).  $\tilde{Q}(\tau)$  é a estatística Ljung-Box calculada utilizando os primeiros  $\tau$  coeficientes. Sob a hipótese nula de que  $\{z_j\}$  é uma série ruído branco,  $\tilde{Q}(\tau) \sim \chi^2(\tau-r)$ , em que  $r$  é o número total de parâmetros do modelo.

No que se refere ao efeito "Páscoa", confirmou-se a presença dois tipos de efeitos "expansivo" e "depressivo". Embora o coeficiente associado ao primeiro efeito (0.0459) tenha vindo superior, em valor absoluto, ao do segundo (0.0302), esperar-se-ia, contudo, que em termos relativos essa discrepância fosse muito mais elevada, ou seja, que o efeito "expansivo" sobre as vendas fosse mais acentuado.

No que respeita ao efeito "dias da semana", confirma-se a ideia de que a existência de um sábado adicional num determinado mês do ano, relativamente ao mês homólogo do ano anterior, terá um impacte bastante forte sobre o indicador que, no entanto, parece ser algo exagerado. Neste caso, a taxa de crescimento do indicador situar-se-á cerca de 5 por cento acima do nível registado numa situação em que, mantendo tudo o resto constante, a variável  $\Delta_{12} Z_{6t}$  assuma o valor zero.

Relativamente aos restantes dias da semana, e ao contrário do que se esperava, as estimativas de alguns coeficientes vieram estatisticamente não significativas<sup>[20]</sup>, pelo que variações homólogas não nulas das variáveis  $Z_{5t}$  e  $Z_{7t}$  não terão qualquer impacte sobre o índice, significando que o nível de vendas registado nas sexta-feiras e domingos tende a aproximar-se do nível médio semanal. Note-se que a detecção deste efeito foi realizada com base em informação sobre as vendas diárias de um hipermercado, num determinado ano. Pode acontecer que as restantes unidades da amostra, sobretudo os supermercados, apresentem um padrão de evolução semanal das suas vendas bastante distinto.

Quanto ao efeito "rendimento", as variáveis de intervenção a ele associadas acabaram por ser eliminadas do modelo. Durante o processo de estimação os coeficientes de  $SEXR_t$  e  $DOMR_t$  vieram sistematicamente com sinal contrário ao esperado, enquanto que o coeficiente da variável  $SABR_t$ , embora com sinal positivo, se revelou estatisticamente não significativo. Na realidade, embora este efeito possa estar bem visível na informação diária, pode acontecer que se dilua totalmente quando essa informação é agregada em termos mensais. Por outro lado, a elevada irregularidade subjacente às taxas de variação homólogas do índice poderá, de certo modo, dificultar a detecção deste efeito.

Por último, no que concerne ao efeito "novo hipermercado", os resultados empíricos da estimação do modelo (4.1) indicam a presença do fenómeno do desvio de comércio em apenas quatro casos, mais concretamente, após a abertura do segundo, quinto, sétimo e

---

<sup>[20]</sup> Para testar a significância estatística de cada parâmetro considerou-se um nível de significância de 5%, o que equivale a tomar 1.67 como valor crítico para o respectivo "t-rácio".

nono hipermercados. Verifica-se que este fenómeno foi mais intenso nos 12 meses que se seguiram à entrada em funcionamento do segundo hipermercado, tendo a taxa de crescimento do índice do consumo corrente, de acordo com a estimativa de  $\eta_2$ , ficado subavaliada durante esse período em cerca de 4.3 por cento.

Relativamente ao efeito "*overshooting*", constata-se que apenas em três casos as taxas de crescimento das vendas de novas unidades parecem ter tido algum impacto sobre a taxa de crescimento do indicador. Repare-se, no entanto, que parece existir algum exagero na magnitude dos coeficientes associados às variáveis de intervenção  $DEN_{1t}$  e  $DEN_{2t}$ , sobretudo dos que respeitam à taxa à qual o efeito "*overshooting*" de cada um dos hipermercados se amortece ao longo do tempo. Em termos concretos, a obtenção de  $\delta_H$  próximas da unidade traduz-se na existência de intervenções cujos efeitos sobre a série assumem um carácter quase permanente. Por exemplo, de acordo com os resultados obtidos, verifica-se que a entrada da taxa de crescimento das vendas do primeiro hipermercado no cálculo do índice ainda exerce, decorridos 6 anos, um impacto de cerca de 1 por cento sobre a taxa do indicador. A existência de raízes quase unitárias nos polinómios  $(1 - \delta_1 B)$  e  $(1 - \delta_2 B)$  poderá ser o reflexo de a série estar "contaminada" por um conjunto de intervenções cujos momentos de ocorrência se localizam muito proximamente uns dos outros.

Como foi referido no capítulo I, a existência de raízes unitárias em pelo menos um dos polinómios  $\delta_m(B)$  poderá induzir a que se sobre-estime erradamente o grau de integrabilidade da série original. Tendo em conta os resultados empíricos que se obtiveram, parece ser este o caso da série em estudo. Na realidade, analisando as figuras 1 e 3, relativas à série do ICC<sub>t</sub> e às respectivas taxas de variação homólogas, e as figuras 22 a 24 referentes às funções de autocorrelação das séries  $LICC_t$ ,  $\Delta_{12}LICC_t$  e  $\Delta\Delta_{12}LICC_t$ , tudo indica que seriam necessárias uma diferença de ordem um e uma diferença de ordem 12 para tornar a série  $LICC_t$  estacionária. Por conseguinte, a série original aparentemente seria  $I(1,1)$ . Contudo, após se ter estimado o modelo (4.1) verificou-se que o padrão de comportamento das funções de autocorrelação e autocorrelação parcial da respectiva série residual era comparável ao de uma série estacionária (figuras 16 e 17, respectivamente), pelo que, de acordo com o modelo, a série  $LICC_t$  poderia ser  $I(0,1)$ . É interessante notar que a transformação que se operou na variável *output* do modelo de intervenção (4.1), tendo em vista a especificação do modelo em termos de variações homólogas, correspondeu, ao fim e ao cabo, à estacionarização da componente residual do modelo, ou seja, da própria variável  $LICC_t$ .

No que se refere à qualidade do ajustamento da parte residual do modelo, pode considerar-se razoável, embora se detecte ainda alguma autocorrelação residual (tabela 1.3.C). Os valores da estatística  $\tilde{Q}$  de Ljung-Box para os desfasamentos 24 e 36 permitem aceitar a hipótese nula de inexistência de autocorrelação apenas um nível de significância de 2.0 e 9.4 por cento, respectivamente. Contudo, todas os coeficientes de autocorrelação da série residual estão dentro das bandas de significância de Bartlett ( $2\sigma$ ).

Relativamente aos "choques" exógenos não captados pelas variáveis de intervenção, foram detectadas nove observações atípicas, o que confirma uma vez mais a elevada irregularidade subjacente às taxas de variação homólogas do índice (tabela 1.3.B). Repare-se na redução substancial do desvio padrão do ajustamento com correcção de *outliers* quando comparado com o do ajustamento sem correcção de observações "anormais".

Com base nas estimativas dos coeficientes das variáveis de intervenção indicadas na tabela 1.3.A, a série das variações homólogas do índice do consumo corrente corrigida dos efeitos das intervenções ( $\Delta_{12} \text{ LICC}_t^*$ ) pode ser obtida da seguinte forma<sup>[21]</sup>:

$$\begin{aligned} \Delta_{12} \text{ LICC}_t^* = & \Delta_{12} \text{ LICC}_t - 0.0158 \Delta_{12} \left( \sum_{j=1}^4 Z_{jt} \right) - 0.0133 \Delta_{12} Z_{5t} - 0.0519 \Delta_{12} Z_{8t} - \\ & - 0.0046 \Delta_{12} Z_{7t} - 0.0459 \Delta_{12} \text{ PASP}_t + 0.0302 \Delta_{12} \text{ PASN}_t + \\ & + 0.0425 \text{ DA}_2 + 0.0178 \text{ DA}_{5t} + 0.0117 \text{ DA}_{7t} + 0.0247 \text{ DA}_{9t} - \\ & - \frac{0.1803}{(1 - 0.9632 B)} \text{ DEN}_{1t} - \frac{0.0974}{(1 - 0.9066 B)} \text{ DEN}_{2t} - 0.0652 \text{ DEN}_{4t} \quad (4.3) \end{aligned}$$

Na figura 25, em anexo, estão representadas as séries das taxas de variação homólogas publicada e corrigida. Da análise desta figura pode-se constatar que as taxas de crescimento corrigidas, embora ligeiramente mais alisadas que as originais, continuam a apresentar um elevado grau de irregularidade. Por outro lado, é notório que o ajustamento efectuado para corrigir os efeitos da entrada dos hipermercados 1 e 2<sup>[22]</sup> no

<sup>[21]</sup> Para efeitos de correcção do indicador resolveu-se manter todas as variáveis associadas ao efeito "dias da semana", embora algumas delas sejam estatisticamente não significativas. Por outro lado, decidiu-se também corrigir retrospectivamente o período 1985:1 a 1985:6.

<sup>[22]</sup> Estas ocorreram em Janeiro de 1987 e Outubro de 1988, respectivamente.



cálculo do índice acabou por dominar por completo o comportamento do mesmo, tendo um impacto que parece excessivamente depressivo.

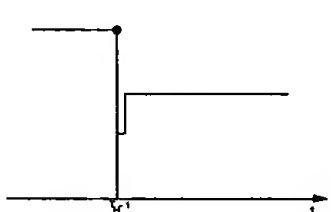
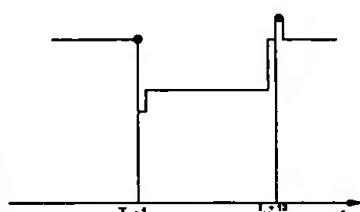
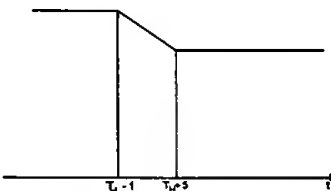
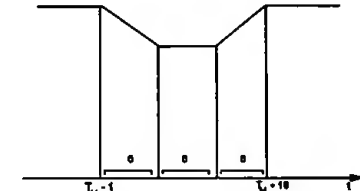
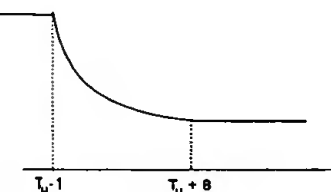
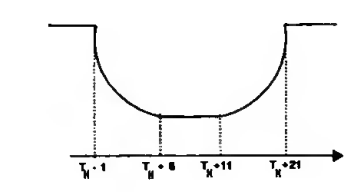
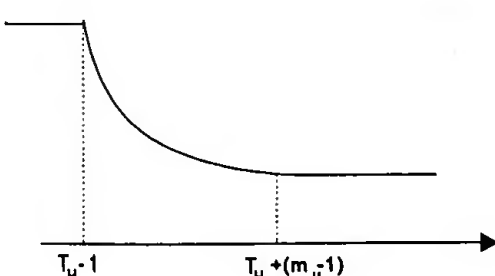
Uma das potenciais limitações do modelo estimado reside na forma utilizada para captar o efeito "desvio de comércio", que pode ser considerada muito restritiva. De facto, além de pressupor que a abertura de uma nova unidade tenderá a subavaliar a taxa do indicador apenas durante 12 meses, tem também subjacente a existência de uma única especificação para modelar o referido efeito em todos os casos de abertura. Tendo como objectivo colmatar estas lacunas foi ensaiado um conjunto de especificações alternativas para captar o efeito "desvio de comércio", cujas representações gráficas surgem incluídas na tabela 1.4.

Mesmo com estas especificações mais realistas e flexíveis os resultados não foram muito satisfatórios, tendo-se a especificação de um modelo adequado para a componente residual revelado bastante difícil. Por outro lado, continuou a verificar-se a existência de raízes quase unitárias em alguns dos polinómios  $(1-\delta_H B)$  das variáveis de intervenção associadas ao efeito "*overshooting*".

Os resultados empíricos obtidos nesta primeira hipótese de correcção do indicador do consumo corrente reflectem claramente a inadequabilidade de algumas variáveis de intervenção para captar o efeito de determinados fenómenos, mais concretamente, dos associados ao aparecimento das grandes superfícies.

Relativamente ao efeito "*overshooting*", ainda que os resultados obtidos tivessem apresentado alguma aderência à realidade, a especificação que se propôs para modelar este efeito tem inerente alguns pontos "fracos". Em primeiro lugar, esta especificação foi construída a partir do comportamento das taxas de crescimento das vendas de um ou dois hipermercados, tendo sido posteriormente generalizada aos restantes. Mesmo que as variáveis de intervenção do efeito "*overshooting*" estejam mal especificadas, os respectivos coeficientes estimados podem vir significativos em resultado de estarem simplesmente a captar ruído da série. Em segundo lugar, a correcção efectuada com base nesta especificação nunca é definitiva, no sentido de que observações adicionais darão origem a novas estimativas dos coeficientes do modelo. Este facto contribuirá, sem dúvida, para uma maior instabilidade das taxas de variação homólogas dos indicador. Por último, em termos prospectivos, tornar-se-á muito complicado corrigir o efeito da entrada de futuros novos hipermercados.

Tabela 1.4 - Impacte da abertura de um novo hipermercado sobre o indicador do consumo corrente

Sobre o nível de vendas	Sobre as taxas do ICC	Observações
		<p>2 - Admite-se que o desvio de comércio é mais acentuado no primeiro mês em que o novo hipermercado abre, após o qual estabiliza a um nível mais baixo.</p>
		<p>3 - O efeito desvio de comércio apenas se completa ao fim de 6 meses.</p>
		<p>4 - O efeito desvio de comércio apenas se completa ao fim de 9 meses.</p>
		<p>5 - O efeito desvio de comércio apenas se completa ao fim de <math>m_H</math> meses, sendo <math>m_H</math> definido como o número de meses ao fim dos quais se detectou uma certa estabilização do nível de vendas do hipermercado H.</p>

### **IV.3 - Segunda Hipótese de Correção do Indicador do Consumo Corrente**

Nesta segunda hipótese pretende-se obter uma série do indicador do consumo corrente na qual o efeito "novo hipermercado" esteja expurgado à partida. Os problemas inerentes à primeira hipótese e os resultados empíricos não muito satisfatórios que nela se obtiveram motivaram a formulação de uma proposta alternativa de correção do indicador em que o efeito "novo hipermercado" pudesse ser eliminado antes de se corrigirem os restantes efeitos.

Em termos concretos, a obtenção de um indicador do consumo corrente intrinsecamente não "contaminado" pelo efeito "novo hipermercado" traduziu-se na reconstrução do próprio índice. A principal diferença entre os dois indicadores, como adiante se verá, assenta no critério de entrada de uma nova unidade no cálculo do índice e na possibilidade de exclusão temporária de unidades afectadas pelo aparecimento de novos supermercados. De facto, a experiência mostrou que alguns dos novos hipermercados apresentaram vendas com taxas de crescimento muito elevadas durante um longo período de tempo<sup>[23]</sup>, pelo que o critério dos "18 meses" revelou-se pouco adequado em grande parte das situações. Além disso, a metodologia de cálculo do indicador "original" não contemplava uma solução para o fenómeno do desvio de comércio, limitação que agora se procurou corrigir logo à partida na fase da sua construção.

#### **IV.3.1 - Construção do "Novo" Indicador do Consumo Corrente**

O "novo" índice foi construído a partir da mesma informação estatística utilizada no cálculo do índice "original", isto é, com base nos valores das vendas nominais de supermercados e hipermercados localizados no país. Neste caso foram apenas consideradas as vendas referentes ao período posterior a Janeiro de 1984, pelo que se tomou essa data como a base do "novo" índice. Com o objectivo de reduzir alguma volatilidade potencialmente introduzida no índice pelas variações dos preços, deflacionaram-se as

---

<sup>[23]</sup> Por exemplo, o primeiro hipermercado que abriu ainda registou taxas de crescimento das suas vendas superiores a 25 por cento (em termos reais), decorridos três anos da sua abertura.

vendas com séries do IPC obtidas de médias móveis centradas de 3 termos das séries do IPC anteriormente utilizadas. Na construção deste "novo" indicador existiu também o cuidado em respeitar a estabilidade da amostra entre dois meses homólogos consecutivos, pelo que se manteve a mesma fórmula de cálculo do índice.

Como amostra inicial consideraram-se as unidades que já estavam em funcionamento em Janeiro de 1984. Admitiu-se como hipótese que, nessa data, as vendas dessas unidades estavam a crescer a taxas aproximadas ou, pelo menos, que a variância das taxas de crescimento das vendas desse conjunto de lojas era reduzida<sup>[24]</sup>. Relativamente às unidades que abriram depois de Janeiro de 1984, o momento da sua inclusão na amostra foi determinado com base em critérios estatísticos.

Considere-se a seguinte especificação para modelar o comportamento das taxas de crescimento homólogas das vendas reais das unidades incluídas na amostra num determinado momento do tempo:

$$t_i = \mu + \varepsilon_i \quad (4.4)$$

em que  $t_i$  é a taxa de crescimento da unidade  $i$ ,  $\mu$  é um parâmetro desconhecido e  $\varepsilon_i$  é uma variável aleatória residual associada a  $t_i$  ( $i=1, \dots, n$ ). Assume-se que  $\varepsilon_i \sim N(0, \sigma^2)$ . Repare-se que esta especificação, num modelo *cross-section*, equivale a admitir que, em cada mês  $t$ , as taxas de crescimento das vendas das diferentes unidades diferem apenas de um ruído, de valor esperado nulo.

O estimador dos mínimos quadrados para  $\mu$  virá dado por:

$$\hat{\mu} = \frac{\sum_n t_i}{n} \quad (4.5)$$

Suponha-se agora que abre uma nova unidade e se admite que a taxa de crescimento das suas vendas verifica a relação

$$t_{n+1} = \mu_{n+1} + \varepsilon_{n+1} \quad (4.6)$$

tal que  $\varepsilon_{n+1} \sim N(0, \sigma^2)$ .

---

<sup>[24]</sup> Refira-se que em Janeiro de 1984 apenas existiam supermercados.

Embora se esteja num contexto *cross-section*, pode-se aplicar o teste de previsão de Chow (*Chow forecasting test*) para ensaiar a hipótese da permanência de estrutura, isto é,  $\mu_{n+1} = \mu$ . Numa situação em que existe apenas uma observação adicional a estatística de teste vem dada por:

$$F = \frac{e_{n+1}' \left(1 + \frac{1}{n}\right)^{-1} e_{n+1}}{\frac{e'e}{n-1}}$$

em que  $e_{n+1}$  é o erro de previsão associado à observação  $(n+1)$ , ou seja,  $e_{n+1} = t_{n+1} - \hat{\mu}$ , e  $(e'e)$  o somatório dos quadrados dos resíduos da regressão (4.4). Sob a hipótese nula de permanência de estrutura  $F$  tem uma distribuição *F-Snedecor* com  $(1, n-1)$  graus de liberdade.

Considerando as especificações (4.4) e (4.6) e tendo presente a relação que existe entre as distribuições *F-Snedecor* e *t-Student*, a estatística de teste vem definida por:

$$t_1 = \frac{t_{n+1} - \hat{\mu}}{\sqrt{\sum_n (t_i - \hat{\mu})^2 \frac{(n+1)}{n(n-1)}}}$$

que, sob a hipótese nula, tem uma distribuição *t-Student* com  $(n-1)$  graus de liberdade.

Com base na estatística  $t_1$  pode-se estabelecer a seguinte condição para definir o momento de entrada de uma nova unidade no cálculo do índice do consumo corrente: inclui-la na amostra quando a respectiva taxa de crescimento das vendas —  $t_{n+1}$  — estiver contida no seguinte intervalo de confiança

$$\left[ \hat{\mu} \pm t_{\frac{\alpha}{2}} \sqrt{\sum_n (t_i - \hat{\mu})^2 \frac{(n+1)}{n(n-1)}} \right] \quad (4.7)$$

onde  $t_{\alpha/2}$  representa o ponto crítico correspondente a um ensaio com nível de significância de  $\alpha\%$ .

Um dos pressupostos básicos do modelo (4.4) é o de que  $V(\varepsilon_i) = \sigma^2$ ,  $\forall i$ . No entanto, conhecendo as características dos dois tipos de unidades incluídos na amostra — supermercados e hipermercados — é pouco verosímil que os resíduos deste modelo sejam homoescedásticos.

Tendo em vista testar a presença de heteroescedasticidade residual no modelo (4.4) começou-se por se admitir o seguinte padrão de heteroescedasticidade:

$$V(\varepsilon_i) = \sigma^2 \frac{1}{VV_i}, \quad i=1, \dots, n \quad (4.8)$$

em que  $VV_i$  é o volume de vendas da unidade  $i$ .

Assumida esta relação de proporcionalidade inversa entre a variância da variável residual e o volume de vendas, testou-se a presença de heteroescedasticidade com base num conjunto de passos de certa forma análogos aos do ensaio de Goldfeld-Quandt.

Em primeiro lugar, as diversas unidades foram ordenadas por ordem decrescente do seu volume de vendas. Nesta fase constatou-se uma relação de 2 para 1 entre as vendas da grande superfície<sup>[25]</sup> que registou o volume de vendas mais baixo e as vendas do supermercado que registou o volume de vendas mais elevado. Em contrapartida, observou-se que era de 5 para 1 a relação entre as vendas desta última unidade e as vendas do supermercado que registou o volume de vendas mais baixo. Tendo em conta estas proporções, optou-se, numa primeira fase, por testar a existência de heteroescedasticidade apenas entre os diferentes supermercados, deixando de fora as grandes superfícies. Para isso, subdividiu-se a "nova" amostra em três partes iguais e eliminaram-se as observações correspondentes à ordem intermédia. Designando por grupo 3 o conjunto das unidades com volume de vendas mais elevado e por grupo 1 o conjunto com vendas mais baixas e admitindo a independência entre estas duas sub-amostras, o teste de homoescedasticidade pode ser efectuado com base na seguinte estatística do rácio de variâncias:

---

<sup>[25]</sup> Refira-se que, neste caso, foi considerado um conceito mais lato de "grande superfície" que abrangeu, para além dos hipermercados, três grandes supermercados cujo volume de vendas era comparável ao dos hipermercados.

$$R = \frac{S_1^2}{S_3^2} = \frac{\frac{\sum_{n_1} (t_i^{(1)} - \bar{t}^{(1)})^2}{(n_1 - 1)}}{\frac{\sum_{n_3} (t_i^{(3)} - \bar{t}^{(3)})^2}{(n_3 - 3)}}$$

onde  $S_1^2$  e  $S_3^2$  são, respectivamente, as variâncias amostrais do primeiro e terceiro grupos,  $t_i^{(j)}$  é a taxa de crescimento das vendas da unidade  $i$  pertencente ao grupo  $j$  e  $\bar{t}^{(j)}$  a média das taxas de crescimento das vendas do grupo  $j$  ( $j=1,3$ ). Quando  $\sigma_3^2 = \sigma_1^2$  a estatística  $R$  tem uma distribuição  $F(n_1-1, n_3-1)$ .

De acordo com o padrão de heteroscedasticidade acima definido, a hipótese a ensaiar pode ser formulada da seguinte forma:

$$H_0: \sigma_1^2 = \sigma_3^2 \quad \text{versus} \quad H_a: \sigma_1^2 > \sigma_3^2$$

Repare-se que a forma como se especificou as hipóteses nula e alternativa e se definiu a estatística  $R$  permite considerar como região crítica, associada a este teste, uma região unilateral direita formada pelo conjunto de pontos localizados à direita de  $F_\alpha(n_1-1, n_3-1)$ , em que  $\alpha$  é a dimensão do teste.

A estatística de teste foi calculada para três momentos do tempo — Novembro de 1991, Agosto de 1992 e Março de 1993 — por forma a que a rejeição ou aceitação da hipótese nula não se fundamentasse em apenas uma "observação". Houve também o cuidado em escolher meses que, em princípio, não estariam "contaminados" por nenhum dos efeitos "dias da semana", "Páscoa" ou "rendimento". Os resultados do teste estão indicados na seguinte tabela:

1º Ensaio para a Detecção de Heteroscedasticidade			
	Novembro de 1991	Agosto de 1992	Março de 1993
R	4.17	1.42	1.67
$F(n_1-1, n_3-1)$	$F(23,23)$	$F(24,24)$	$F(25,25)$
$\alpha\%$	Valor Crítico	Valor Crítico	Valor Crítico
1%	2.72	2.58	2.60
5%	2.01	1.98	1.96
10%	1.72	1.70	1.68

De acordo com os resultados obtidos, verificou-se que hipótese nula foi rejeitada em apenas um dos casos (Novembro de 1991)<sup>[26]</sup>, pelo que se concluiu, então, que a especificação inicialmente proposta para caracterizar o padrão de heteroscedasticidade talvez não fosse a mais adequada. Consequentemente, propôs-se uma segunda hipótese em que se admitiu que a variabilidade das taxas de crescimento das vendas das grandes superfícies é menor que a variabilidade das taxas de crescimento das vendas das restantes unidades. Designando o primeiro grupo de unidades por grupo 2 e o segundo por 1, esta hipótese equivale a estabelecer a seguinte relação:

$$\sigma_2^2 = k \sigma_1^2 \quad \text{para } 0 < k < 1$$

em que  $\sigma_1^2$  e  $\sigma_2^2$  são, respectivamente, as variâncias das taxas de crescimento das vendas dos grupos 1 e 2.

Ensaia-se a hipótese  $H_0: \sigma_1^2 = \sigma_2^2$  versus  $H_1: \sigma_1^2 > \sigma_2^2$  para os três momentos do tempo considerados no primeiro teste com base numa estatística R equivalente.

<sup>[26]</sup> Refira-se, contudo, que se obteve um resultado "perverso" para Agosto de 1992, em que a estatística de teste, referente a esta data, assumiu um valor inferior à unidade. Deste modo, o valor apresentado na tabela corresponde a  $1/R$ . Mesmo assim, verificou-se a não rejeição da hipótese nula para este caso.



2º Ensaio para a Detecção de Heteroscedasticidade			
	Novembro de 1991	Agosto de 1992	Março de 1993
R	1.59	5.56	5.88
$F(n_1-1, n_2-1)$	$F(70,6)$	$F(75,7)$	$F(79,9)$
$\alpha\%$	Valor Crítico	Valor Crítico	Valor Crítico
1%	7.03	9.17	4.44
5%	3.73	4.42	2.77
10%	2.76	3.13	2.20

Conforme se pode constatar da tabela acima apresentada, a hipótese nula foi rejeitada em dois dos três casos<sup>[27]</sup>, pelo que se considerou válida a hipótese de que a variância das taxas de crescimento das vendas das grandes superfícies é menor do que a variância das taxas de crescimento das restantes unidades.

Em conformidade com estes resultados, definiu-se a seguinte condição como critério de entrada da taxa de crescimento das vendas de uma nova unidade no cálculo do indicador: a unidade (n+1) entra no cálculo do índice sse

$$t_{n+1} \in \left[ \hat{\mu} \pm t_{\frac{\alpha}{2}} \sqrt{\frac{\sum_{n_1} (t_i^{(1)} - \hat{\mu})^2 + \frac{1}{k} \sum_{n_2} (t_i^{(2)} - \hat{\mu})^2}{n-1}} k \left( 1 + \frac{1}{k n_1 + n_2} \right) \right] \quad (4.9)$$

ou

$$t_{n+1} \in \left[ \hat{\mu} \pm t_{\frac{\alpha}{2}} \sqrt{\frac{\sum_{n_1} (t_i^{(1)} - \hat{\mu})^2 + \frac{1}{k} \sum_{n_2} (t_i^{(2)} - \hat{\mu})^2}{n-1}} \left( 1 + \frac{k}{k n_1 + n_2} \right) \right] \quad (4.10)$$

<sup>[27]</sup> Embora, para Agosto de 1992, a hipótese nula não tenha sido rejeitada a 1%, optou-se, mesmo assim, por rejeitá-la uma vez que, para os outros dois níveis de significância, os resultados apontaram nesse sentido.

consoante se trate de uma grande superfície ou de um supermercado de menor dimensão, respectivamente<sup>[28]</sup>. Nas expressões anteriores,  $k$  corresponde ao rácio  $S_2^2/S_1^2$ .

Repare-se que a condição de entrada inicialmente apresentada para o modelo homoescedástico não é mais do que um caso particular das condições acima apresentadas, pois, para  $k=1$ , as expressões (4.9) e (4.10) reduzem-se à expressão (4.7).

Refira-se ainda que as regiões de confiança, construídas para determinar o momento da inclusão de uma nova unidade no cálculo do índice, podem também ser utilizadas para detectar casos de desvio de comércio. Como se referiu na primeira hipótese de correcção, a abertura de uma nova unidade tende, em geral, a originar um desvio das vendas das unidades já existentes para essa nova unidade<sup>[29]</sup>. Consequentemente, as unidades afectadas pelo fenómeno do desvio de comércio tenderão a apresentar, durante alguns meses, uma quebra nas taxas de variação homólogas das suas vendas. No caso de estas unidades já estarem incluídas no cálculo do índice será conveniente que sejam eliminadas da amostra enquanto durar o desvio do comércio, por forma a não enviesarem para baixo a taxa de crescimento do indicador. Suponhamos uma situação em que se verifica a abertura de uma nova grande superfície e se pretende testar se uma determinada unidade está afectada pelo efeito do "desvio de comércio". Suponha-se também que, nesse momento do tempo, o indicador está a ser calculado com base em  $(n+1)$  unidades e que a unidade potencialmente afectada é a  $n$ -ésima mais uma. A decisão de excluir esta unidade pode ser determinada pela condição de  $t_{n+1}$  não pertencer ao intervalo definido pela expressão (4.9) ou (4.10), consoante se trate de uma grande superfície ou de um supermercado, respectivamente.

Com base neste conjunto de critérios estatísticos, definidos para determinar o momento da entrada de uma nova unidade na amostra, bem como a exclusão de unidades afectadas pelo efeito "desvio do comércio", obteve-se um "novo" índice do consumo corrente

---

<sup>[28]</sup> A dedução destes resultados encontra-se no apêndice apresentado no final do trabalho. Note-se, todavia, desde já, que o estimador de  $\mu$  já não vem dado por uma média simples das taxas de crescimento das vendas das unidades incluídas na amostra, antes corresponde ao estimador GLS do modelo com heteroescedasticidade.

<sup>[29]</sup> Regra geral, as unidades mais afectadas tenderão a ser as que se localizam nas proximidades da nova unidade.

que se designou ICC2<sup>[30]</sup>. Na figura 28, em anexo, comparam-se as taxas de variação homólogas do "novo" índice com as publicadas, ou seja, do ICC. Como se pode constatar, é bastante notória a diferença de comportamentos das duas séries no período compreendido entre 1987 a 1990, altura em que o índice "original" apresentou taxas de crescimento anormalmente elevadas. Como foi referido na primeira hipótese, este comportamento das taxas de variação homólogas do ICC deveu-se basicamente à entrada no cálculo do índice do primeiro hipermercado que abriu, cujas vendas cresceram a taxas muito altas durante bastante tempo. Relativamente ao grau de volatilidade subjacente às duas séries verifica-se que a variância da série "original" é superior ao dobro da variância da "nova" série, como aliás se pode inferir da simples inspecção gráfica.

#### IV.3.2 - Correção dos Efeitos "Dias da Semana", "Páscoa" e "Rendimento"

Uma vez que o "novo" índice foi expurgado apenas do efeito "novo hipermercado" (efeito "desvio de comércio" e efeito "*overshooting*") estará, naturalmente, também afectado pelos outros três efeitos, ou seja, pelo efeito "dias da semana", efeito "Páscoa" e efeito "rendimento". Como a natureza da "nova" série do indicador do consumo corrente é semelhante à da série original (ambas são construídas a partir da mesma informação estatística), foram consideradas, para o indicador ICC2, as mesmas variáveis de intervenção previamente definidas na primeira hipótese para o indicador ICC. Desta forma, especificou-se o seguinte modelo de intervenção:

$$\begin{aligned} \text{LICC2}_t = & \alpha_0 + \sum_{j=1}^7 \alpha_j Z_{jt} + \varphi_1 \text{PASP}_t + \varphi_2 \text{PASN}_t + \\ & + \lambda_1 \text{SEXR}_t + \lambda_2 \text{SABR}_t + \lambda_3 \text{DOMR}_t + N_t \end{aligned} \quad (4.11)$$

em que LICC2<sub>t</sub> é o logaritmo do "novo" índice do consumo corrente (ICC2<sub>t</sub>).

Note-se que no modelo (4.11) foi considerada a especificação não restrita para captar o efeito "dias da semana" uma vez que, com a eliminação das variáveis de

---

<sup>[30]</sup> Na figura 26, em anexo, está representada graficamente a série do ICC2. Os valores desta série vêm indicados na coluna 2 da tabela 2.2.

intervenção associadas ao efeito "novo hipermercado", deixou de existir o problema do reduzido número de graus de liberdade (o número de parâmetros associados a estas variáveis de intervenção era 19!).

Tendo-se estimado o modelo (4.11) com base no período 1984:7 a 1993:3 (105 observações) obteve-se uma série residual não estacionária. Como se pode observar da análise da figura 27, em anexo, a forte sazonalidade subjacente à série do logaritmo do "novo" índice indicia a existência de uma raiz unitária na componente sazonal. A corroborar este resultado está a função de autocorrelação da série LICC2<sub>t</sub> (figura 29), cujos picos correspondentes à componente sazonal, para além de elevados, tendem lentamente para zero. Para estacionarizar a série LICC2<sub>t</sub> recorreu-se a uma diferença de ordem 12 —  $\Delta_{12} = (1 - B^{12})$  LICC2<sub>t</sub> —, pelo que o modelo (4.11) foi re-especificado da seguinte forma:

$$\begin{aligned} \Delta_{12} \text{LICC2}_t = & \alpha_0 + \sum_{j=1}^7 \alpha_j \Delta_{12} Z_{jt} + \varphi_1 \Delta_{12} \text{PASP}_t + \varphi_2 \Delta_{12} \text{PASN}_t + \\ & + \lambda_1 \Delta_{12} \text{SEXR}_t + \lambda_2 \Delta_{12} \text{SABR}_t + \lambda_3 \Delta_{12} \text{DOMR}_t + \zeta_t \end{aligned} \quad (4.12)$$

Com o objectivo de obter estimadores mais eficientes para os coeficientes das variáveis de intervenção, especificou-se um modelo ARMA(1,0) para a componente residual, ainda que as funções de autocorrelação e autocorrelação parcial da primeira estimativa da série residual do modelo (4.12) — figuras 31 e 32, respectivamente — sejam compatíveis com um comportamento de quase ruído branco por parte desta variável. Posteriormente, procedeu-se à estimação e avaliação do diagnóstico do modelo de intervenção de acordo com o procedimento descrito na secção 2.1.2 do capítulo I, tendo-se obtido os resultados indicados nas tabelas 2.3.A a 2.3.C<sup>[31]</sup>.

Relativamente ao efeito "dias da semana" verifica-se que, conforme seria de esperar, todos os coeficientes associados a este efeito são estatisticamente significativos. Novamente se constata que a existência de um sábado adicional, num determinado mês, terá um inpace bastante forte sobre a taxa de crescimento do índice. No que se refere aos restantes dias da semana, verifica-se que a magnitude das estimativas dos coeficientes das

---

<sup>[31]</sup> A matriz das correlações dos coeficientes estimados encontra-se em anexo na tabela 2.3.D.

segundas, quartas, sextas e domingos é aproximadamente igual, o que significa que o efeito associado a estes dias tenderá a ser idêntico. Este resultado vem, de certa forma, opôr-se à evidência empírica reflectida nas figuras 4 a 7, mas convém ter presente que estas reflectem apenas o comportamento de um único hipermercado durante algumas semanas de um dado ano.

No que respeita ao efeito "Páscoa", os coeficientes estimados de ambas as variáveis de intervenção —  $\hat{\phi}_1$  e  $\hat{\phi}_2$  — vieram estatisticamente significativos, tendo-se constatado, tal como se esperava, que o acréscimo habitualmente registado nas vendas em virtude da ocorrência da Páscoa (efeito "expansivo") mais que compensa a contracção que tende a ocorrer nas vendas durante esse mesmo período (efeito "depressivo").

Relativamente às variáveis de intervenção do efeito "rendimento", estas acabaram por ser eliminadas do modelo pelas mesmas razões mencionadas na primeira hipótese de correcção do indicador. Em consequência da eliminação destas variáveis, houve necessidade de melhorar a especificação para a componente residual, tendo-se retido na especificação "final" um modelo ARMA(1,7), cujos parâmetros constam também da tabela 2.3.A.

A avaliação do diagnóstico do modelo ARMA referente à série residual indica que, em termos gerais, esta especificação é adequada para modelar a componente não sistemática do modelo de intervenção. A hipótese nula do teste de Ljung-Box é aceite a um nível de significância de 7%, considerando quer 24 quer 36 desfasamentos, e todos os coeficientes de autocorrelação e de autocorrelação parcial estão contidos nos limites definidos pelas bandas de significância de Bartlett e de Daniels-Jenkins, (figuras 34 e 35, respectivamente). Nesta segunda hipótese de correcção considerou-se também um valor crítico de 2.7 para a detecção de *outliers*, tendo sido detectadas seis observações atípicas. Repare-se que, relativamente à primeira hipótese, este número representa uma redução de um terço do número de observações "anormais" detectadas pelo programa. Quanto ao desvio padrão do ajustamento com e sem correcção de *outliers* constata-se que, neste caso, a redução não é assim tão substancial como a verificada na primeira hipótese, facto compreensível uma vez que as taxas de variação homólogas do "novo" índice são menos irregulares que as taxas publicadas<sup>[32]</sup>.

---

<sup>[32]</sup> Veja-se a figura 28.

Tabela 2.3.A - Especificação "Final" do Modelo (4.12)				
	Parâmetro	Variável	Estimativa	t-rácio
1	$\alpha_0$		0.0574	15.73
2	$\alpha_1$	$\Delta_{12}Z_{1t}$	0.0287	4.39
3	$\alpha_2$	$\Delta_{12}Z_{2t}$	0.0185	2.52
4	$\alpha_3$	$\Delta_{12}Z_{3t}$	0.0257	3.65
5	$\alpha_4$	$\Delta_{12}Z_{4t}$	0.0339	4.79
6	$\alpha_5$	$\Delta_{12}Z_{5t}$	0.0269	3.55
7	$\alpha_6$	$\Delta_{12}Z_{6t}$	0.0657	10.41
8	$\alpha_7$	$\Delta_{12}Z_{7t}$	0.0254	3.45
9	$\varphi_1$	$\Delta_{12}PASP_t$	0.0739	9.88
10	$\varphi_2$	$\Delta_{12}PASN_t$	-0.0426	- 5.18
11	$\phi_1$		0.312	2.87
12	$\theta_6$		-0.371	- 3.79
13	$\theta_7$		0.49	4.98
S.E. (sem ajustamento de outliers) = 0.026				
S.E. (com ajustamento de outliers) = 0.017				
Modelo da componente residual: $(1 - \phi_1 B)\zeta_t = (1 - \theta_6 B^6 - \theta_7 B^7)\varepsilon_t$				

Tabela 2.3.B - Detecção e Ajustamento de Outliers			
Momento de Ocorrência	Estimativa	t-rácio	Tipo de outlier
28	-0.031	-4.23	LS
42	0.052	3.09	IO
61	0.085	5.11	IO
85	0.056	6.87	AO
89	0.073	4.38	IO
102	-0.026	-3.04	LS

Tabela 2.3.C - Coeficientes de Autocorrelação da Série Residual

$\tau$	1	2	3	4	5	6	7	8	9	10	11	12
$\rho(\tau)$	.01	.08	-.02	-.08	-.06	-.07	.07	.09	-.07	.01	0.4	-.08
$\tilde{Q}(\tau)$	0	.7	.7	1.3	1.6	2.1	2.6	3.5	4.1	4.1	4.2	4.9
$\tau$	13	14	15	16	17	18	19	0	21	22	23	24
$\rho(\tau)$	-.06	-.10	.10	.05	.01	-.12	.10	.08	-.09	.21	.04	.02
$\tilde{Q}(\tau)$	5.3	6.4	7.5	7.7	7.8	9.4	10.6	11.4	12.4	18.0	18.2	18.2
$\tau$	25	26	27	28	29	30	31	32	33	34	35	36
$\rho(\tau)$	.04	-.03	-.15	-.10	-.12	.02	-.02	-.15	-.18	0	.02	-.08
$\tilde{Q}(\tau)$	18.5	18.6	21.5	22.9	24.7	24.7	24.8	27.9	32.5	32.5	32.6	33.6

As variações homólogas do "novo" índice de consumo corrente corrigidas dos efeitos "dias da semana" e "Páscoa" —  $\Delta_{12} \text{LICC2}_t^*$  — podem ser obtidas do cálculo da seguinte expressão:

$$\begin{aligned}
 \Delta_{12} \text{LICC2}_t^* = & \Delta_{12} \text{LICC2}_t - 0.0287 \Delta_{12} Z_{1t} - 0.0185 \Delta_{12} Z_{2t} - 0.0257 \Delta_{12} Z_{3t} - \\
 & - 0.0339 \Delta_{12} Z_{4t} - 0.0269 \Delta_{12} Z_{5t} - 0.0657 \Delta_{12} Z_{6t} - \\
 & - 0.0254 \Delta_{12} Z_{7t} - 0.0739 \Delta_{12} \text{PASP}_t + 0.0426 \Delta_{12} \text{PASN}_t
 \end{aligned} \quad (4.13)$$

Na figura 37 e nas colunas 3 e 4 da tabela 2.1 comparam-se as taxas de crescimento homólogas não corrigidas e as corrigidas do "novo" índice. Como se pode constatar, a existência de alterações na composição do calendário tende a ter um forte impacto sobre a série das taxas de variação homólogas do indicador do consumo corrente. Embora subsista alguma irregularidade na série das taxas, talvez devida à própria natureza da série, verifica-se que a série corrigida é muito mais alisada. Por outro lado, comparando esta série com a série obtida na hipótese 1 — figura 38 — confirma-se que, como foi mencionado anteriormente, a correcção do efeito "novo hipermercado" efectuada na primeira hipótese imprimiu um carácter excessivamente depressivo sobre as taxas de variação homólogas do índice. Esta figura permite ainda observar que a série das taxas corrigidas na hipótese 1 é bastante mais irregular que a obtida na hipótese 2.

No modelo (4.12), ao contrário do modelo (4.1), considerou-se a especificação não restrita para modelar o efeito "dias da semana". Tendo em vista testar a validade da especificação que se admitiu na primeira hipótese de correcção para modelar este efeito, estimou-se o seguinte modelo:

$$\Delta_{12} \text{ LICC2}_t = \alpha_0 + \sum_{j=1}^7 \alpha_j \Delta_{12} Z_{jt} + \varphi_1 \Delta_{12} \text{PASP}_t + \varphi_2 \Delta_{12} \text{PASN}_t + \zeta_t \quad (4.14)$$

impondo as seguintes restrições:  $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4$ .

Os resultados da estimação constam na tabela 2.4. Com base nos somatórios dos quadrados dos resíduos do modelo restrito e não restrito, testou-se a validade da hipótese  $H_0: \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4$ . O valor da estatística F indica que a hipótese nula é rejeitada para um nível de significância de 3.5 por cento, o que significa que a condição imposta na primeira hipótese de correcção, embora fosse considerada admissível, surge como pouco verosímil.

Tabela 2.4	Parâmetro	Variável	Estimativa	t-rácio
1	$\alpha_0$		0.0611	18.78
2	$\alpha_1$ $\alpha_2$ $\alpha_3$ $\alpha_4$	$\Delta_{12}Z_{1t}$ $\Delta_{12}Z_{2t}$ $\Delta_{12}Z_{3t}$ $\Delta_{12}Z_{4t}$	0.0207	2.99
3	$\alpha_5$	$\Delta_{12}Z_{5t}$	0.0265	3.40
4	$\alpha_6$	$\Delta_{12}Z_{6t}$	0.062	9.13
5	$\alpha_7$	$\Delta_{12}Z_{7t}$	0.0203	2.76
6	$\varphi_1$	$\Delta_{12}\text{PASP}_t$	0.0777	11.15
7	$\varphi_2$	$\Delta_{12}\text{PASN}_t$	-0.0357	- 4.58
8	$\phi_1$		0.356	3.46
9	$\theta_6$		-0.256	- 2.60
10	$\theta_7$		0.492	5.18



Tabela 2.4 - (cont.)
S.E. (sem ajustamento de <i>outliers</i> ) = 0.027 S.E. (com ajustamento de <i>outliers</i> ) = 0.017
Modelo da componente residual: $(1 - \phi_1 B)\zeta_t = (1 - \theta_6 B^6 - \theta_7 B^7)\varepsilon_t$
$H_0: \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4$ $F(3,79) = 3.0$ Nível de significância = 0.035

Nesta segunda hipótese de correcção estimou-se também o modelo (4.12) admitindo para o efeito "dias da semana" a re-especificação (3.2) apresentada na secção 1.1 do capítulo III:

$$\begin{aligned} \Delta_{12}\text{LICC2}_t = & \alpha_0 + \sum_{j=1}^6 \beta_j \Delta_{12}(Z_{jt} - Z_{7t}) + \beta_7 (\Delta_{12}(\sum_{j=1}^7 Z_{jt}) + \\ & + \varphi_1 \Delta_{12}\text{PASP}_t + \varphi_2 \Delta_{12}\text{PASN}_t + \zeta_t \end{aligned} \quad (4.15)$$

em que  $\beta_j = (\alpha_j - \bar{\alpha})$ ,  $\beta_7 = \sum_{j=1}^7 \alpha_j / 7 = \bar{\alpha}$  e  $\sum_{j=1}^7 Z_{jt}$  mede o número total de dias do mês  $t$ .

Note-se que, neste caso, o coeficiente  $\beta_7$  vai captar os efeitos decorrentes da existência de anos bissextos e não propriamente o efeito "dias do mês", uma vez que este último efeito se anula em termos homólogos excepto para alguns meses de Fevereiro. Como se pode observar da análise da tabela 2.5.A, este coeficiente veio estatisticamente significativo indicando que a taxa de crescimento das vendas num mês de Fevereiro bissexto tenderá a ser superior, em cerca de 3.4 pontos percentuais, à do mês homólogo do ano precedente. Verifica-se também que algumas das estimativas dos coeficientes "dias da semana" vieram estatisticamente não significativas. Este resultado apenas traduz a existência de parâmetros  $\alpha_j$ , mais concretamente  $\alpha_1$ ,  $\alpha_4$  e  $\alpha_6$ , cuja magnitude não se afasta muito da média dos sete coeficientes diários. Em contrapartida, as estimativas dos coeficientes das variáveis  $(Z_{2t} - Z_{7t})$  e  $(Z_{6t} - Z_{7t})$  associadas, respectivamente, à terça-feira e ao sábado evidenciam um maior afastamento relativamente à média. Os resultados obtidos

permitem ainda confirmar que ao definirem-se as variáveis do efeito "dias da semana" como o desvio entre o número de dias do tipo  $j$  e o número de domingos do mês  $t$  ( $j$ =segunda, terça,...,sábado) consegue-se obter uma maior ortogonalidade entre estas variáveis (tabela 2.5.B).

Tabela 2.5.A	Parâmetro	Variável	Estimativa	t-rácio
1	$\alpha_0$		0.0574	15.73
2	$\beta_1$	$\Delta_{12}(Z_{1t} - Z_{7t})$	-0.0034	- 1.33
3	$\beta_2$	$\Delta_{12}(Z_{2t} - Z_{7t})$	-0.0136	- 5.13
4	$\beta_3$	$\Delta_{12}(Z_{3t} - Z_{7t})$	-0.064	- 2.77
5	$\beta_4$	$\Delta_{12}(Z_{4t} - Z_{7t})$	0.0018	0.73
6	$\beta_5$	$\Delta_{12}(Z_{5t} - Z_{7t})$	-0.0052	- 1.92
7	$\beta_6$	$\Delta_{12}(Z_{6t} - Z_{7t})$	0.0336	13.34
8	$\beta_7$	$\Delta_{12} \sum_{j=1}^7 Z_{jt}$	0.0321	4.87
9	$\varphi_1$	$\Delta_{12} \text{PASP}_t$	0.0739	9.88
10	$\varphi_2$	$\Delta_{12} \text{PASN}_t$	-0.0426	- 5.18
11	$\phi_1$		0.312	2.87
12	$\theta_6$		-0.371	- 3.79
13	$\theta_7$		0.49	4.98
S.E. (sem ajustamento de outliers) = 0.026				
S.E. (com ajustamento de outliers) = 0.017				
Modelo da componente residual: $(1 - \phi_1 B) = (1 - \theta_6 B^6 - \theta_7 B^7)$				

Tabela 2.5.B - Coeficientes de Correlação das Estimativas dos Parâmetros													
	1	2	3	4	5	6	7	8	9	10	11	12	13
1	1												
2	0	1											
3	0	-.66	1										
4	0	.07	-.39	1									
5	.01	.18	-.21	-.42	1								
6	0	-.30	.17	-.02	-.54	1							
7	0	.28	-.12	-.06	.24	-.73	1						
8	0	-.25	.15	.04	.04	.18	-.31	1					
9	-.01	.27	-.12	-.27	-.14	.34	-.17	-.11	1				
10	.01	-.37	.36	.14	.04	-.24	.14	.14	-.87	1			
11	.01	-.08	.04	.14	.03	-.01	-.12	.11	-.40	.34	1		
12	.01	-.04	.05	-.07	-.02	-.04	.06	-.06	-.06	.06	.04	1	
13	.03	-.01	0	.02	-.04	.07	-.05	.10	-.04	.03	.07	-.01	1

Relativamente à modelação do efeito "Páscoa", foi também considerada a especificação apresentada por Escrivá e Sastre, na qual as variáveis  $PASP_t$  e  $PASN_t$  foram substituídas por duas novas variáveis de intervenção que se designaram  $PASP_t^*$  e  $PASN_t^*$ . Estas variáveis foram construídas a partir dos mesmos coeficientes de ponderação diários obtidos para os efeitos "expansivo" e "depressivo" (tabelas 1.1 e 1.2, respectivamente), tendo-se decidido imputar a totalidade de cada um dos efeitos associados ao efeito "Páscoa" (efeito "expansivo" e "depressivo") ao mês de Abril. Com base na definição da estrutura da variável do efeito "Páscoa" indicada no capítulo III obtiveram-se os seguintes valores para os meses de Março e Abril<sup>[33]</sup>:

<sup>[33]</sup> Nos restantes meses estas variáveis assumem o valor zero.

Tabela 2.6				
Anos	Variável $PASP_t^*$		Variável $PASN_t^*$	
	Março	Abril	Março	Abril
1985	0.20	-0.20	0	0
1986	1.00	-1.00	0.79	-0.79
1987	0	0	0	0
1988	0.83	-0.83	0	0
1989	1.0	-1.0	1.00	-1.00
1990	0	0	0	0
1991	1.00	-1.00	0.67	-0.67
1992	0	0	0	0
1993	0	-	0	-

A estimação do modelo (4.12) com esta especificação alternativa para o "efeito" Páscoa" conduziu exactamente aos mesmos resultados que os apresentados nas tabelas 2.3.A a 2.3.C. Estes resultados estão em conformidade com os esperados, uma vez que a definição das variáveis do modelo de intervenção (4.12) em termos de variações homólogas implica que  $\Delta_{12} PASP_t = \Delta_{12} PASP_t^*$  e  $\Delta_{12} PASN_t = \Delta_{12} PASN_t^*$ . Deste modo, concluiu-se que, embora a especificação proposta por Escrivá e Sastre seja aconselhável quando o objectivo final é corrigir o nível da série<sup>[34]</sup>, ela torna-se perfeitamente indiferente, relativamente à especificação que se propôs inicialmente, nas situações em que o objectivo final é obter uma série de taxas de variação homólogas corrigidas.

Refira-se, por último, que nesta segunda hipótese procedeu-se ainda à correcção dos efeitos "dias da semana" e "Páscoa" pelo procedimento X11ARIMA<sup>[35]</sup>. Tendo em vista comparar os resultados deste procedimento automático com os obtidos na metodologia da

<sup>[34]</sup> Na medida em que este não é alterado quando se corrige a série do efeito "Páscoa".

<sup>[35]</sup> Na figura 39, em anexo, estão representadas as taxas de variação homólogas do "novo" índice não corrigidas e as corrigidas pelo método X11ARIMA.

análise de intervenção, representam-se na figura 40 as séries das taxas de variação homólogas do "novo" índice corrigidas com base em cada uma das metodologias. Como se pode verificar, a série ajustada pelo modelo de intervenção é muito mais alisada que a obtida no procedimento X11ARIMA. Para além disso, existem ainda indícios, na série obtida do X11ARIMA, de que o efeito "Páscoa" não ficou totalmente corrigido, atendendo a que persistem alguns pontos anómalos localizados nos meses de Março e Abril. Daqui se conclui que a flexibilidade inerente à metodologia da análise de intervenção potencia a obtenção de melhores resultados na correcção de efeitos sistemáticos presentes em séries temporais, como sejam o efeito "dias da semana" e o efeito "Páscoa".

## V - CONSIDERAÇÕES FINAIS

Os resultados obtidos nas aplicações empíricas desenvolvidas neste trabalho, sobre o indicador do consumo corrente, permitem-nos tecer algumas considerações sobre a metodologia apresentada para modelar séries temporais afectadas por fenómenos de carácter exógeno e determinístico.

A análise de intervenção, complementada pela análise de *outliers*, ao preencher importantes lacunas dos modelos ARIMA univariados, constitui uma extensão de grande relevo e utilidade, em termos de aplicados, da metodologia univariada Box-Jenkins. De facto, ao permitir expurgar do andamento das séries temporais efeitos determinísticos que interferem com o seu comportamento, nele introduzindo uma variabilidade artificial, esta metodologia revela-se merecedora de um lugar de destaque, no contexto das metodologias de tratamento de séries temporais, além de que, uma grande parte das séries temporais, utilizadas em estudos empíricos, tenderá a estar afectada por efeitos deste tipo.

Todavia, constata-se que, em determinados casos particulares, a simplicidade de que, teoricamente, se reveste a metodologia da análise de intervenção pode não ter correspondência em termos aplicados. Por exemplo, em séries temporais "contaminadas" por múltiplas intervenções que ocorrem bastante próximas umas das outras, como é o caso da série do índice do consumo corrente considerada na primeira hipótese de correcção, torna-se complicado identificar, numa base individual, os efeitos de cada uma delas, dada a sua sobreposição no tempo. Por conseguinte, nestes casos, em que o número de parâmetros do modelo tende a ser elevado, torna-se difícil propôr uma especificação inicial para a função transferência associada a cada variável de intervenção que seja simultaneamente acertada e parcimoniosa, além de que, só em muito raras situações se possuirá informação *a priori* que permita ter uma ideia inicial bastante acertada sobre os efeitos das intervenções. Por outro lado, em situações em que a série temporal está afectada por intervenções cujos efeitos não se esgotaram durante o período amostral utilizado para efeitos de estimação, esta metodologia não oferece uma solução explícita. Neste caso, e como foi referido anteriormente, a especificação "final" encontrada nunca será definitiva, no sentido de que observações adicionais darão origem a novas estimativas dos coeficientes do modelo.

## APÊNDICE: Definição do Critério de Entrada das Grandes Superfícies no Cálculo do Índice do Consumo Corrente

Neste apêndice deriva-se a estatística de teste com base na qual se define o critério de entrada da taxa de crescimento das vendas de uma nova grande superfície no cálculo do indicador do consumo corrente.

Esta estatística foi calculada partindo do pressuposto de que a variância das taxas de crescimento das vendas das grandes superfícies é menor do que a variância das taxas de crescimento das vendas das unidades de menor dimensão, ou seja, dos supermercados. Designando o grupo das grandes superfícies por grupo 2 e o das restantes unidades por 1, ter-se-á a seguinte relação:

$$\sigma_2^2 = k \sigma_1^2 \Leftrightarrow k = \frac{\sigma_2^2}{\sigma_1^2} \quad \text{tal que } 0 < k < 1 \quad (\text{A.1})$$

em que  $\sigma_1^2$  e  $\sigma_2^2$  são, respectivamente, as variâncias das taxas de crescimento das vendas das unidades dos grupos 1 e 2.

Considere-se a seguinte especificação para modelar a taxa de crescimento homóloga das vendas reais da unidade  $i$  —  $t_i$  — num determinado momento do tempo:

$$t_i = \mu + \varepsilon_i \quad (\text{A.2})$$

em que  $\mu$  é um parâmetro desconhecido e  $\varepsilon_i$  é uma variável aleatória residual associada à taxa de crescimento  $t_i$  ( $i=1, \dots, n$ ).

Como os resíduos do modelo (A.2) são heteroescedásticos, em que o padrão de heteroescedasticidade dado pela relação (A.1), admitem-se as seguintes distribuições para os vectores  $\varepsilon_i^{(1)}$  e  $\varepsilon_i^{(2)}$ :

$$\begin{cases} \varepsilon_i^{(1)} \sim N(0, \sigma_1^2) \\ \varepsilon_i^{(2)} \sim N(0, \sigma_2^2) \end{cases}$$

em que  $\varepsilon^{(1)}$  é o vector dos resíduos associado às taxas de crescimento das vendas das unidades de menor dimensão e  $\varepsilon^{(2)}$  o vector dos resíduos associado às taxas de crescimento das vendas das grandes superfícies<sup>[1]</sup>.

A obtenção de um estimador *blue* (*best linear unbiased estimator*) para o parâmetro  $\mu$  implica transformar as variáveis por forma a que os resíduos do modelo transformado sejam homoescedásticos. É fácil verificar que o modelo transformado vem dado por

$$\begin{cases} t_i^{(1)} = \mu + \varepsilon_i^{(1)} & i = 1, \dots, n_1 \\ \frac{t_i^{(2)}}{\sqrt{k}} = \frac{\mu}{\sqrt{k}} + \frac{\varepsilon_i^{(2)}}{\sqrt{k}} & i = 1, \dots, n_2 \end{cases}$$

pois  $V \left( \frac{\varepsilon_i^{(2)}}{\sqrt{k}} \right) = \sigma_1^2$ .

---

<sup>[1]</sup>  $\varepsilon^{(1)}$  e  $\varepsilon^{(2)}$  têm, respectivamente, dimensão  $(n_1 \times 1)$  e  $(n_2 \times 1)$ , com  $n_1 + n_2 = n$ .



Matricialmente, este modelo pode escrever-se na forma

$$\begin{bmatrix} t^{(1)} \\ \vdots \\ \frac{t^{(2)}}{\sqrt{k}} \end{bmatrix} = \begin{bmatrix} 1 \\ \vdots \\ \frac{1}{\sqrt{k}} \end{bmatrix} \mu + \begin{bmatrix} \varepsilon^{(1)} \\ \vdots \\ \frac{\varepsilon^{(2)}}{\sqrt{k}} \end{bmatrix} = \begin{bmatrix} 1 \\ \vdots \\ \frac{1}{\sqrt{k}} \end{bmatrix} \mu + \varepsilon^* \quad (\text{A.3})$$

em que  $\text{Var}(\varepsilon^*) = \sigma_1^2 I$  ( $I$  é a matriz identidade de dimensão  $(n \times n)$ ).

Estimando o modelo (A.3) pelo método dos mínimos quadrados obtém-se o seguinte estimador para  $\mu$ :

$$\begin{aligned} \hat{\mu} &= \left[ \begin{bmatrix} 1 & | & \frac{1}{\sqrt{k}} \end{bmatrix} \begin{bmatrix} 1 \\ \vdots \\ \frac{1}{\sqrt{k}} \end{bmatrix} \right]^{-1} \begin{bmatrix} 1 & | & \frac{1}{\sqrt{k}} \end{bmatrix} \begin{bmatrix} t^{(1)} \\ \vdots \\ \frac{t^{(2)}}{\sqrt{k}} \end{bmatrix} = \\ &= \left( n_1 + \frac{n_2}{k} \right)^{-1} \left( \sum_{n_1} t_i^{(1)} + \frac{1}{k} \sum_{n_2} t_i^{(2)} \right) = \\ &= \frac{\sum_{n_1} t_i^{(1)} + \frac{1}{k} \sum_{n_2} t_i^{(2)}}{\frac{k n_1 + n_2}{k}} = \frac{k \sum_{n_1} t_i^{(1)} + \sum_{n_2} t_i^{(2)}}{k n_1 + n_2} \quad (\text{A.4}) \end{aligned}$$

Admita-se agora que se verifica a abertura de uma nova grande superfície cuja taxa de crescimento é  $t_{n+1}$ . Sob a hipótese de permanência de estrutura, ter-se-á:

$$t_{n+1} = \mu + \varepsilon_{n+1} \quad (\text{A.5})$$

tal que  $E(\varepsilon_{n+1})=0$  e  $V(\varepsilon_{n+1})=k \sigma_1^2$ .

O erro de previsão associado à observação (n+1) vem dado por:

$$\begin{aligned}
 e_{n+1} &= \frac{t_{n+1}}{\sqrt{k}} - \frac{\hat{\mu}}{\sqrt{k}} = \frac{\mu}{\sqrt{k}} + \frac{\varepsilon_{n+1}}{\sqrt{k}} - \frac{1}{\sqrt{k}} \left( \frac{\sum_{n_1} t_i^{(1)} + \frac{1}{\sqrt{k}} \sum_{n_2} t_i^{(2)}}{n_1 + \frac{n_2}{k}} \right) = \\
 &= \frac{\mu}{\sqrt{k}} + \frac{\varepsilon_{n+1}}{\sqrt{k}} - \frac{1}{\sqrt{k} n_1 + \frac{n_2}{\sqrt{k}}} \left( \sum_{n_1} (\mu + \varepsilon_i^{(1)}) + \frac{1}{\sqrt{k}} \sum_{n_2} \frac{t_i^{(2)}}{\sqrt{k}} \right) = \\
 &= \frac{\mu}{\sqrt{k}} + \frac{\varepsilon_{n+1}}{\sqrt{k}} - \frac{1}{\sqrt{k} n_1 + \frac{n_2}{\sqrt{k}}} \left[ n_1 \mu + \sum_{n_1} \varepsilon_i^{(1)} + \frac{1}{\sqrt{k}} \sum_{n_2} \left( \frac{\mu}{\sqrt{k}} + \frac{\varepsilon_i^{(2)}}{\sqrt{k}} \right) \right] = \\
 &= \frac{\mu}{\sqrt{k}} + \frac{\varepsilon_{n+1}}{\sqrt{k}} - \frac{1}{\sqrt{k} n_1 + \frac{n_2}{\sqrt{k}}} \left( n_1 \mu + \sum_{n_1} \varepsilon_i^{(1)} + \frac{n_2 \mu}{k} + \frac{1}{\sqrt{k}} \sum_{n_2} \frac{\varepsilon_i^{(2)}}{\sqrt{k}} \right) = \\
 &= \frac{\mu}{\sqrt{k}} + \frac{\varepsilon_{n+1}}{\sqrt{k}} - \frac{1}{\sqrt{k} n_1 + \frac{n_2}{\sqrt{k}}} \left( n_1 + \frac{n_2}{k} \right) \mu - \frac{1}{\sqrt{k} n_1 + \frac{n_2}{\sqrt{k}}} \left( \sum_{n_1} \varepsilon_i^{(1)} + \frac{1}{\sqrt{k}} \sum_{n_2} \frac{\varepsilon_i^{(2)}}{\sqrt{k}} \right) = \\
 &= \frac{\varepsilon_{n+1}}{\sqrt{k}} - \left( \frac{\sum_{n_1} \varepsilon_i^{(1)} + \frac{1}{\sqrt{k}} \sum_{n_2} \frac{\varepsilon_i^{(2)}}{\sqrt{k}}}{n_1 + \frac{n_2}{k}} \right) = \frac{1}{\sqrt{k}} \left( \varepsilon_{n+1} - \frac{\sum_{n_1} \varepsilon_i^{(1)} + \frac{1}{\sqrt{k}} \sum_{n_2} \frac{\varepsilon_i^{(2)}}{\sqrt{k}}}{n_1 + \frac{n_2}{k}} \right)
 \end{aligned}$$

O valor esperado e a variância do erro de previsão são, respectivamente, iguais a:

$$E(e_{n+1}) = 0$$

$$\begin{aligned}
V(e_0) &= \frac{1}{k} k \sigma_1^2 + \left( \frac{1}{n_1 + \frac{n_2}{k}} \right)^2 \left( n_1 \sigma_1^2 + \frac{1}{k} n_2 \sigma_1^2 \right) = \sigma_1^2 + \frac{1}{k} \frac{1}{n_1 + \frac{n_2}{k}} \sigma_1^2 = \\
&= \left( 1 + \frac{1}{k} \frac{1}{n_1 + \frac{n_2}{k}} \right) \sigma_1^2 = \left( 1 + \frac{1}{k} \frac{k}{k n_1 + n_2} \right) \sigma_1^2 = \\
&= \left( 1 + \frac{1}{k n_1 + n_2} \right) \sigma_1^2
\end{aligned}$$

A estatística de teste do ensaio da permanência de estrutura efectuado na secção 3.1 do capítulo IV vem, neste caso, dada por

$$t_2 = \frac{\frac{1}{\sqrt{k}} (t_{n+1} - \hat{\mu})}{\sqrt{\frac{\sum_{n_1} (t_i^{(1)} - \hat{\mu})^2 + \frac{1}{k} \sum_{n_2} (t_i^{(2)} - \hat{\mu})^2}{n-1}} \sqrt{1 + \frac{1}{k n_1 + n_2}}}$$

ou seja

$$t_2 = \frac{t_{n+1} - \hat{\mu}}{\sqrt{\frac{\sum_{n_1} (t_i^{(1)} - \hat{\mu})^2 + \frac{1}{k} \sum_{n_2} (t_i^{(2)} - \hat{\mu})^2}{n-1}} \sqrt{k \left( 1 + \frac{1}{k n_1 + n_2} \right)}} \quad (A.6)$$

que, sob a hipótese nula, tem uma distribuição *t-Student* com  $(n-1)$  graus de liberdade.

Tendo por base a estatística  $t_2$  pode-se definir o critério de entrada da nova grande superfície, no cálculo do índice, do seguinte modo: as vendas da nova unidade só serão

incluídas na amostra, para efeitos de cálculo do índice, quando a respectiva taxa de crescimento estiver contida no seguinte intervalo de confiança

$$t_{n+1} \in \left[ \hat{\mu} \pm t_{\frac{\alpha}{2}} \sqrt{\frac{\sum_{n_1} (t_i^{(1)} - \hat{\mu})^2 + \frac{1}{k} \sum_{n_2} (t_i^{(2)} - \hat{\mu})^2}{n-1} k \left( 1 + \frac{1}{k n_1 + n_2} \right)} \right] \quad (A.7)$$

onde  $t_{\alpha/2}$  representa o ponto crítico correspondente a um ensaio de dimensão  $\alpha\%$ .

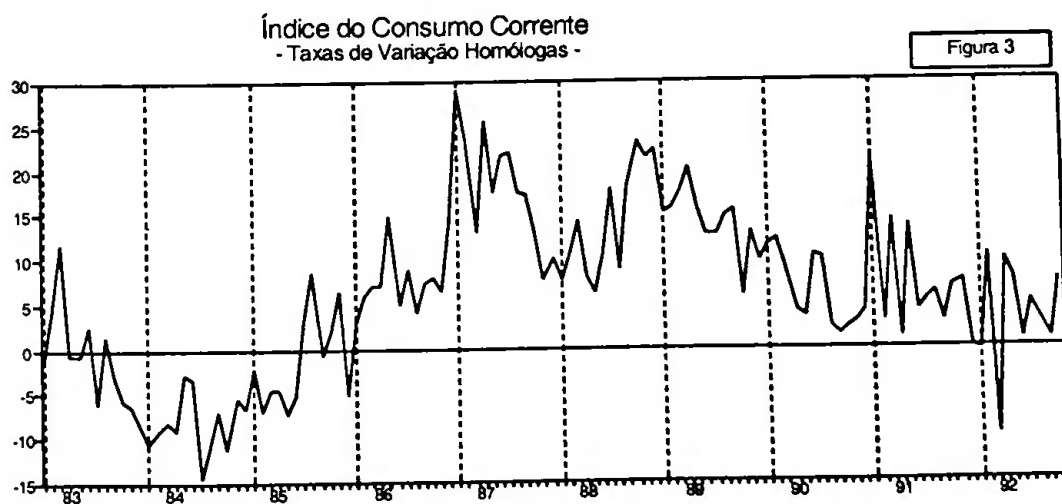
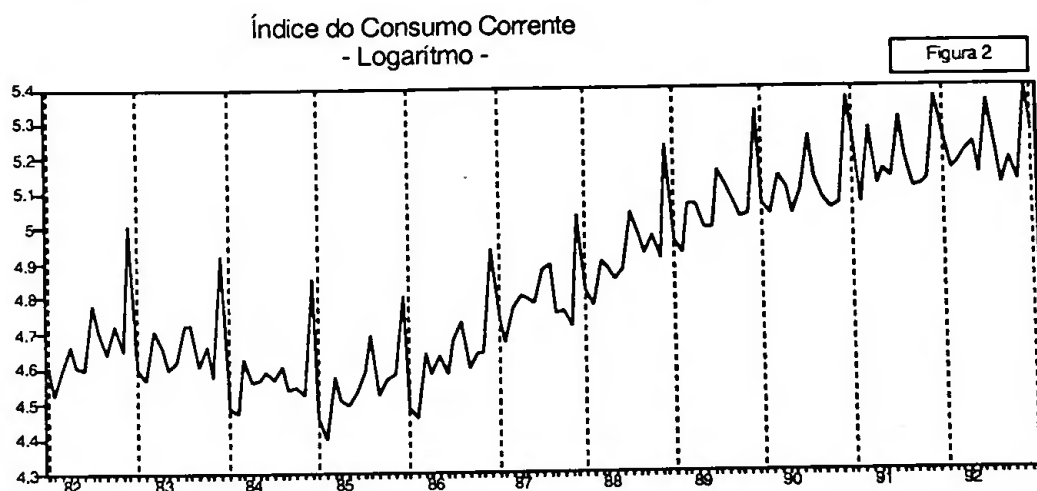
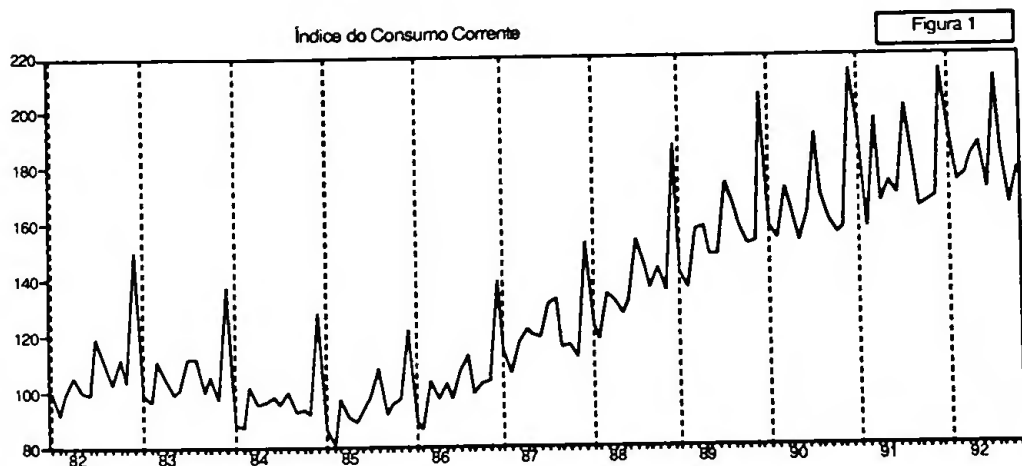
Refira-se ainda que, no caso da nova unidade ser um supermercado<sup>[2]</sup>, o critério de entrada desta unidade no cálculo do indicador vem dado por

$$se \ t_{n+1} \in \left[ \hat{\mu} \pm t_{\frac{\alpha}{2}} \sqrt{\frac{\sum_{n_1} (t_i^{(1)} - \hat{\mu})^2 + \frac{1}{k} \sum_{n_2} (t_i^{(2)} - \hat{\mu})^2}{n-1} \left( 1 + \frac{k}{k n_1 + n_2} \right)} \right] \quad (A.8)$$

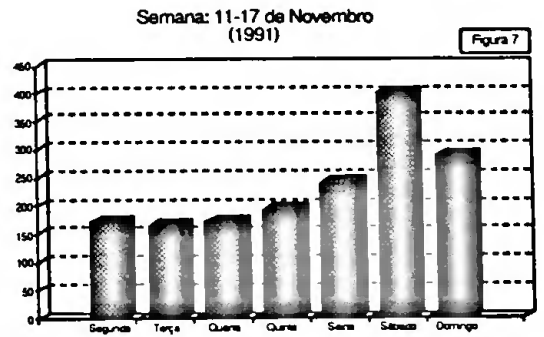
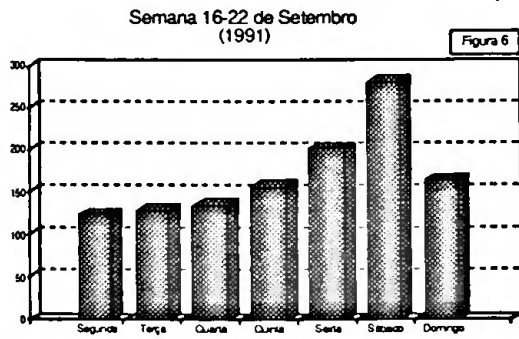
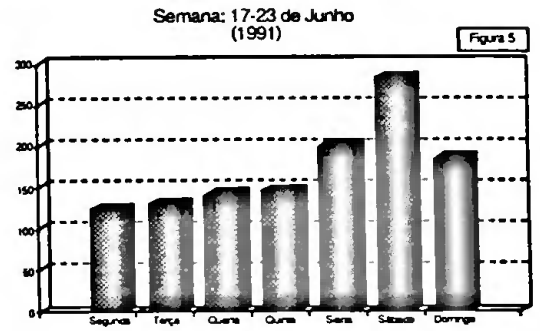
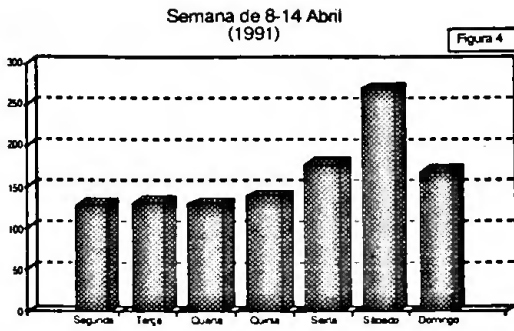
<sup>[2]</sup> Nesta situação, sob a hipótese de permanência de estrutura,  $t_{n+1} = \mu + \varepsilon_{n+1}$ , tal que  $E(\varepsilon_{n+1})=0$  e  $V(\varepsilon_{n+1})=\sigma_1^2$ .

**Anexo de Figuras e Tabelas**

**Primeira Hipótese de Correção**



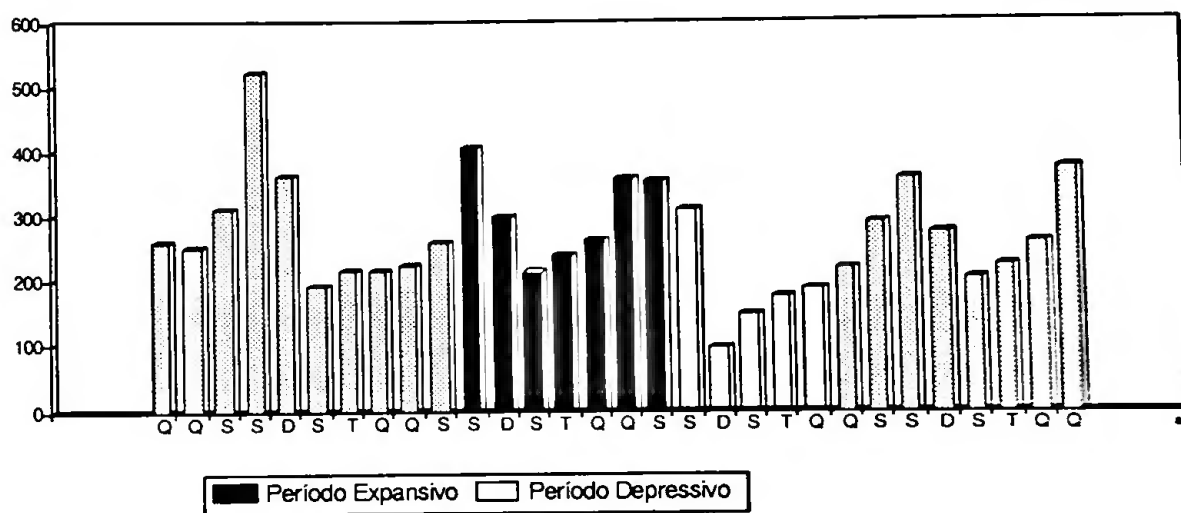
## EFEITO "DIAS DA SEMANA"



# EFEITO "PÁSCOA"

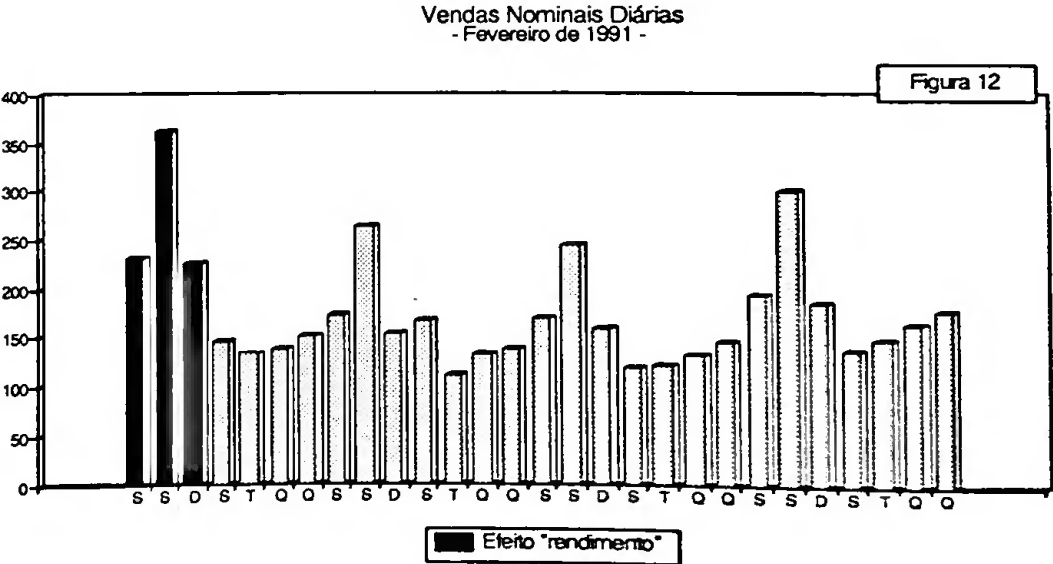
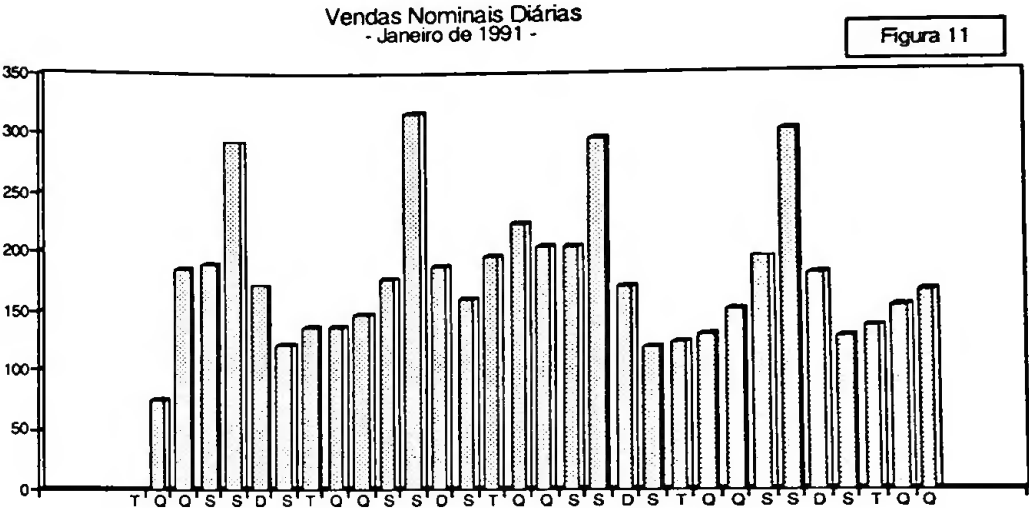
Vendas Nominais Diárias  
- Abril de 1992 -

Figura 8





# EFEITO "RENDIMENTO"





♦ Determinação do Efeito "Rendimento"

Sendo

SEXR= sexta-feira de "rendimento"

SABR= sábado de "rendimento"

DOMR= domingo de "rendimento"

F = feriado

AF = dia que antecede o feriado

Para um mês  $t$  com 31 dias, podem-se ter quatro situações distintas:

	Dia 28	Dia 29	Dia 30	Dia 31
1ª		SEXR	SABR	DOMR
2ª			SEXR	SABR
3ª				SEXR
4ª				

1ª	"fim de semana de rendimento" no mês $t$
----	--

2ª	Dia 28	Dia 29	Dia 30	Dia 31	
			SEXR	SABR	
(i)	F				"fim de semana de rendimento" no mês $t$
(ii)	AF	F			"fim de semana de rendimento" no mês $t$
(iii)		AF	F		"fim de semana de rendimento" no mês $t$
(iv)			AF	F	domingo de "rendimento" no mês $t+1$
(v)				AF	domingo de "rendimento" no mês $t+1$

3ª	Dia 28	Dia 29	Dia 30	Dia 31	
				SEXR	
(i)	F				domingo de "rendimento" no mês t+1
(ii)	AF	F			"fim de semana de rendimento" no mês t
(iii)		AF	F		"fim de semana de rendimento" no mês t
(iv)			AF	F	domingo de "rendimento" no mês t+1
(v)				AF	sábado e domingo de "rendimento" no mês t+1

4ª	Dia 28	Dia 29	Dia 30	Dia 31	
(i)	F				sexta-feira e domingo de "rendimento" no mês t+1
(ii)	AF	F			sábado de "rendimento" no mês t+1
(iii)		AF	F		domingo de "rendimento" no mês t+1
(iv)			AF	F	domingo de "rendimento" no mês t+1
(v)				AF	sábado e domingo de "rendimento" no mês t+1

# SÉRIE RESIDUAL DO MODELO (4.1)

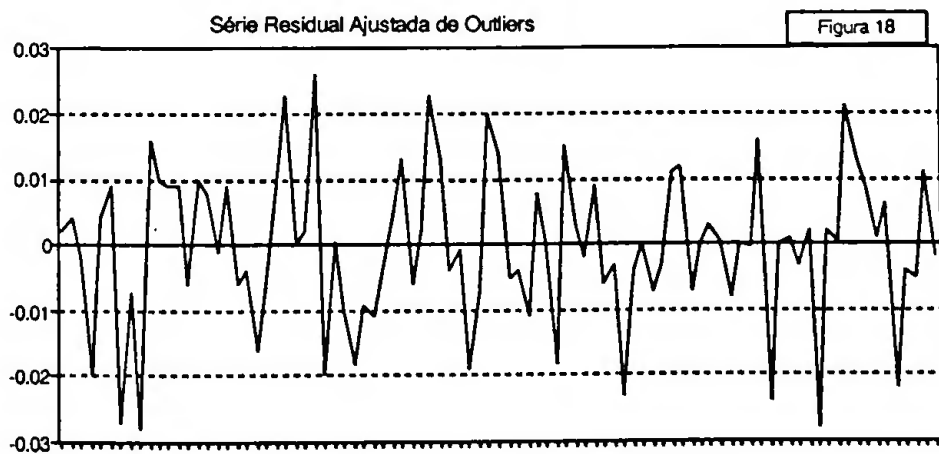
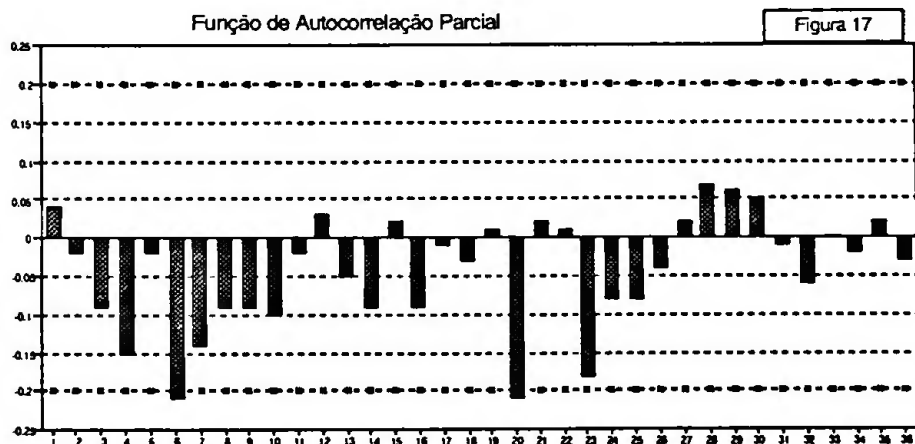
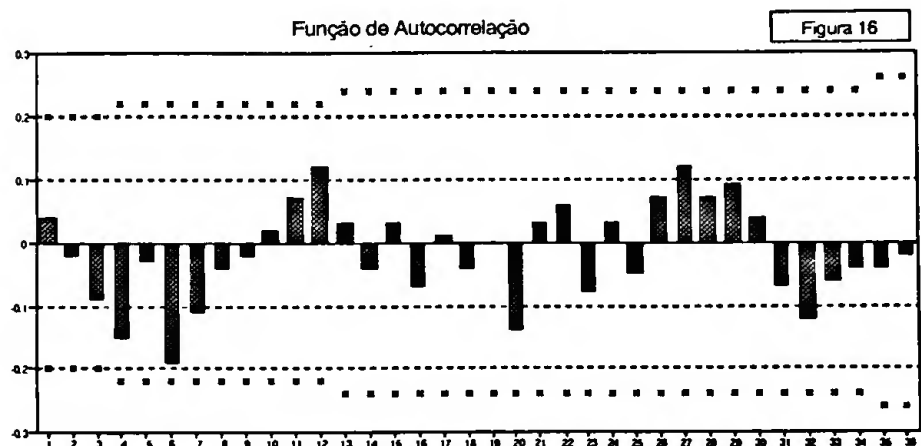
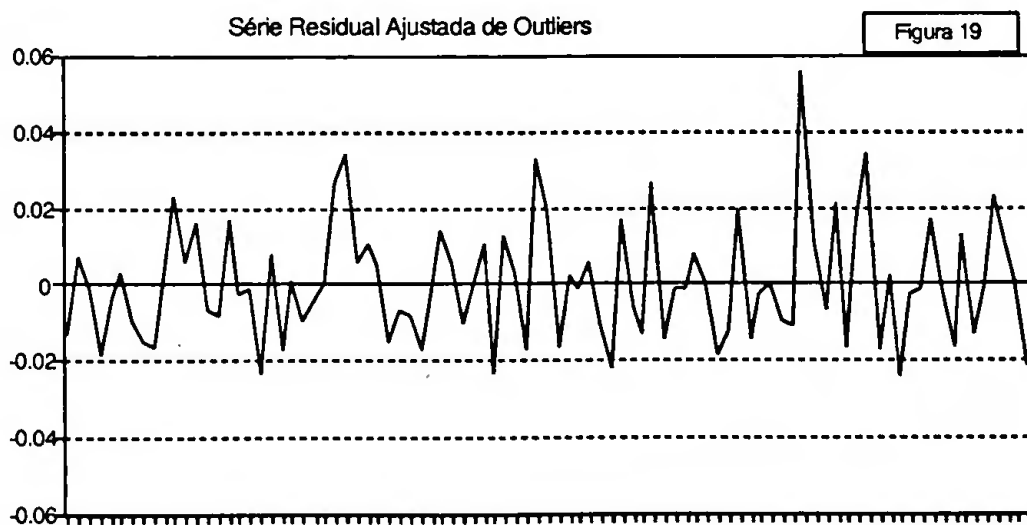
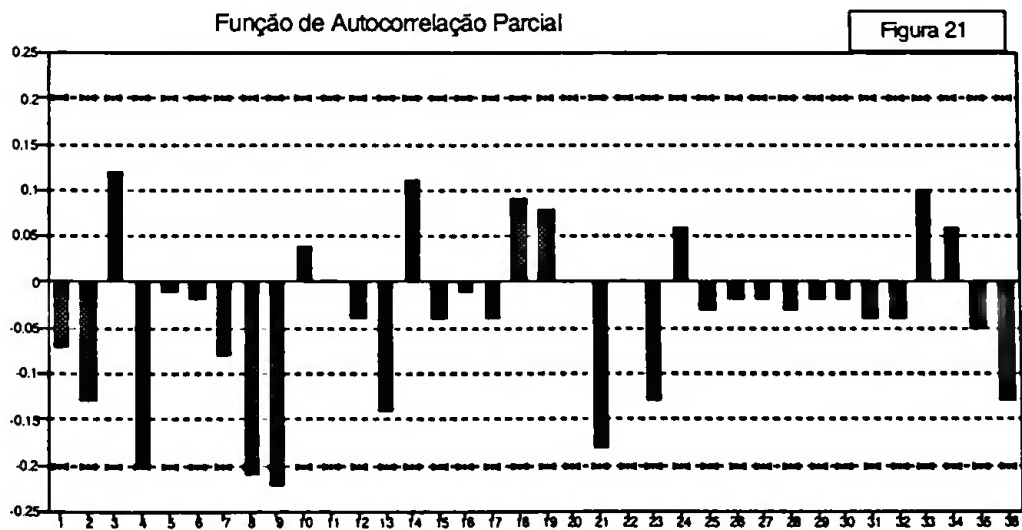
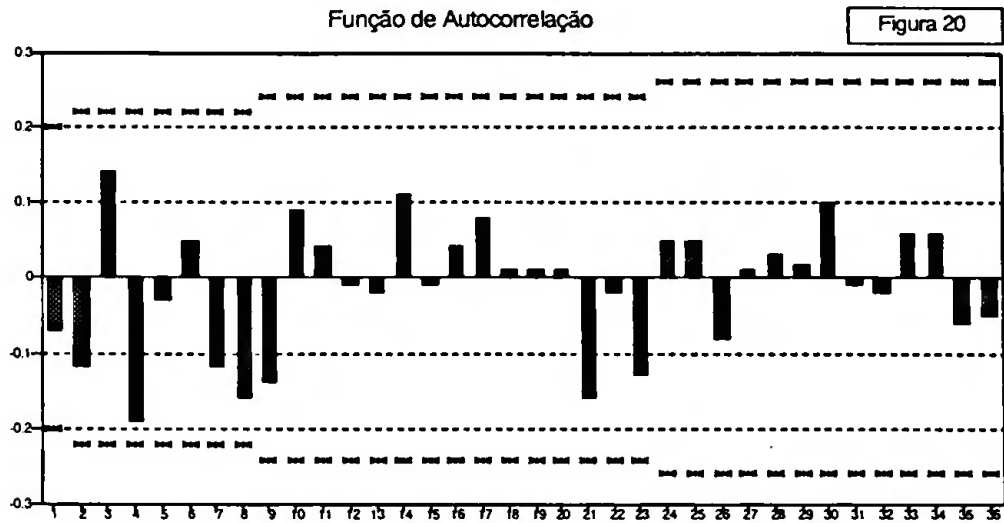


Tabela 1.3.D - Coeficientes de Correlação das Estimativas dos Parâmetros

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	1																
2	-.03	1															
3	-.02	.96	1														
4	-.03	.96	.88	1													
5	-.01	.96	.96	.89	1												
6	0	.07	.16	.06	.09	1											
7	.01	-.08	-.20	-.05	-.12	-.85	1										
8	-.49	.10	.09	.10	.08	0	-.02	1									
9	-.26	-.13	-.13	-.12	-.11	.08	-.01	-.18	1								
10	.17	-.06	-.06	-.06	-.07	-.10	.04	-.36	-.67	1							
11	.12	.25	.26	.24	.22	-.02	-.06	-.03	-.68	.54	1						
12	-.01	.03	.02	.02	.03	-.07	.07	.17	-.65	.51	.05	1					
13	.02	-.01	-.07	-.01	-.03	-.08	.11	-.03	.06	-.04	-.06	-.02	1				
14	-.04	.10	.10	.11	.07	-.06	-.07	.13	-.45	.32	.62	-.14	-.13	1			
15	-.15	-.01	0	-.01	-.03	-.06	-.02	.26	-.65	.48	.61	.29	-.27	.50	1		
16	-.18	.10	.09	.09	.07	-.06	0	.24	-.47	.33	.40	.29	-.01	.34	.40	1	
17	-.06	.15	.21	.13	.17	.23	-.23	.09	-.11	.06	.10	.07	-.11	.07	.11	.15	1

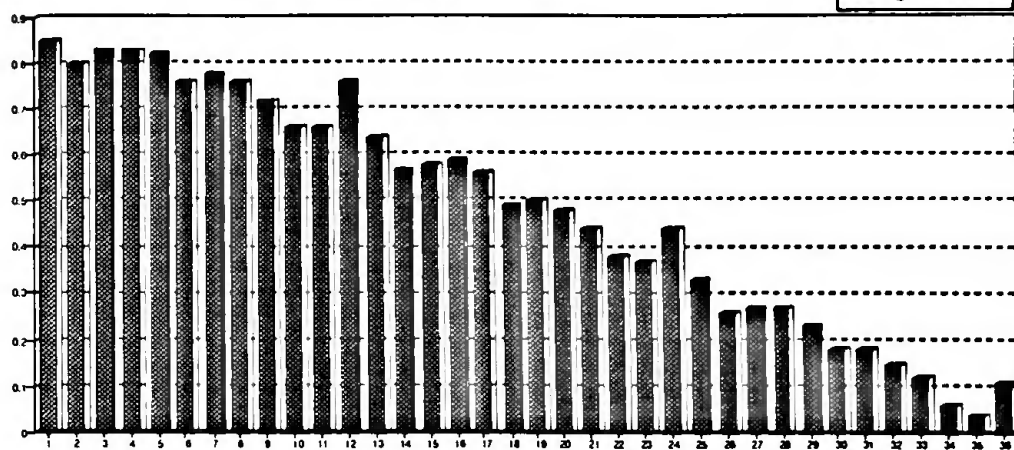


## SÉRIE RESIDUAL DO MODELO "FINAL"



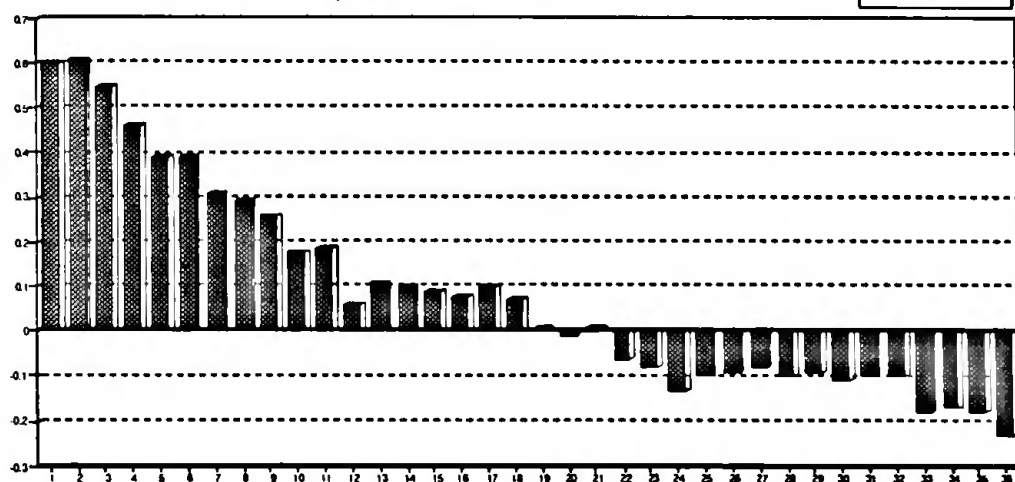
Função de Autocorrelação do LICC

Figura 22



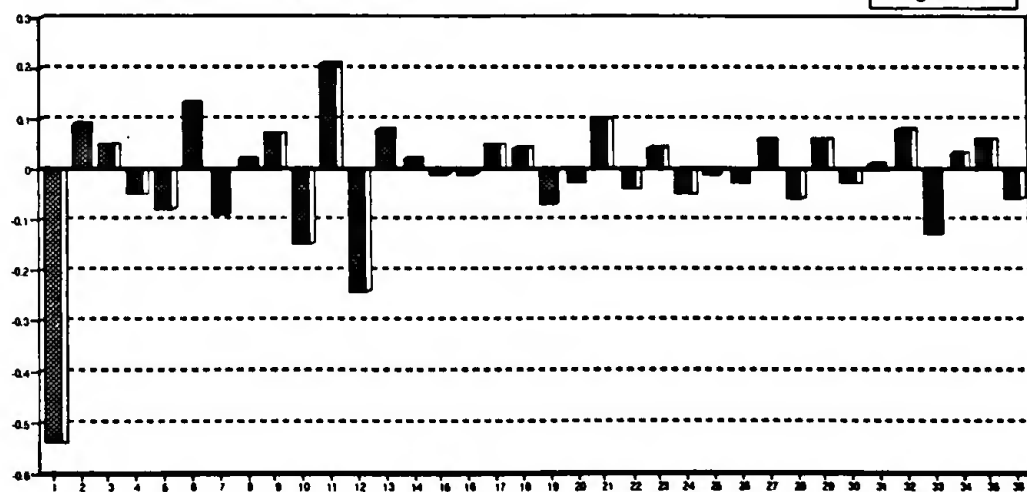
Função de Autocorrelação de  $(1-B^{12})$  LICC

Figura 23



Função de Autocorrelação de  $(1-B)(1-B^{12})$  LICC

Figura 24





Taxas de Variação Homólogas do ICC-Tvh  
- Hipótese 1 -

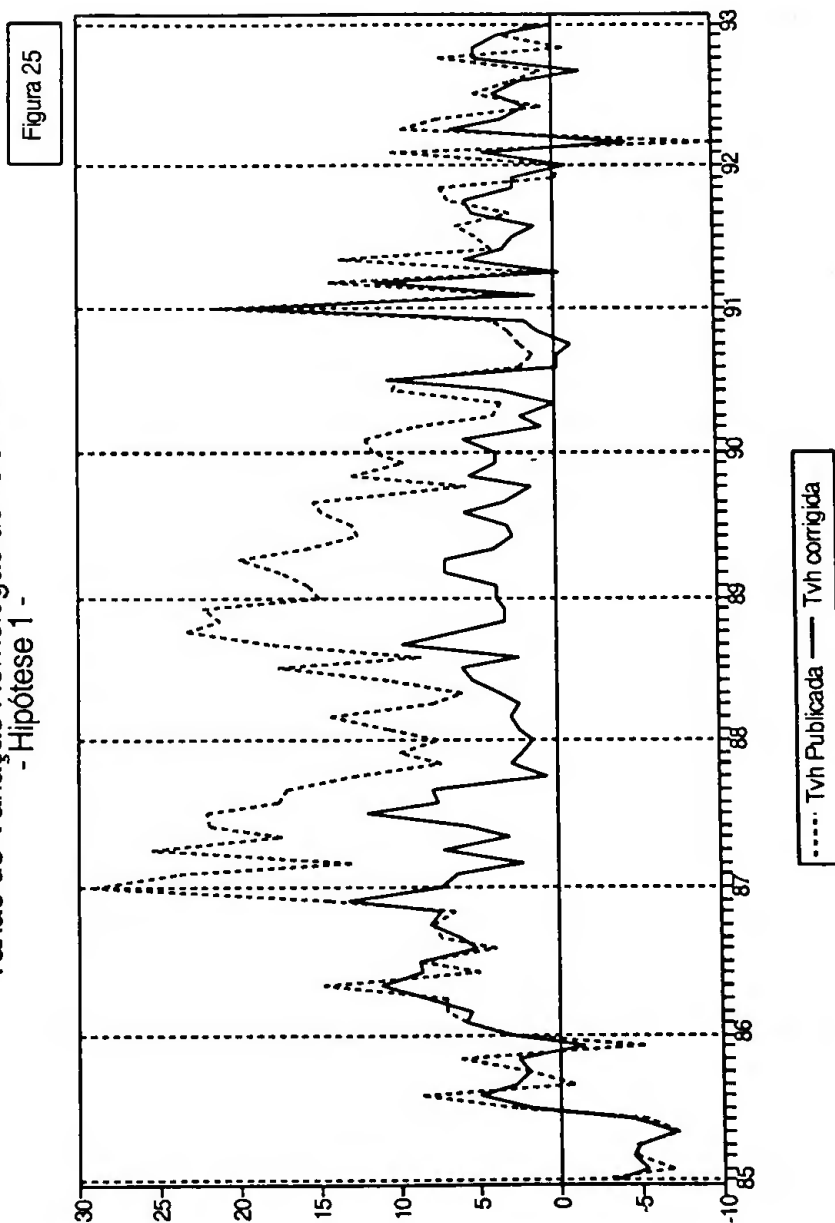


Tabela 1.5 - Indicador do Consumo Corrente (Hipótese 1)

Datas	Índice (Base=1982)		Taxas de variação homólogas	
	Publicado (1)	Corrigido	Publicadas	Corrigidas
1984M1	88.4	88.4		
1984M2	87.5	87.5		
1984M3	101.9	101.9		
1984M4	95.4	95.4		
1984M5	96.4	96.4		
1984M6	98.2	98.2		
1984M7	96.0	96.0		
1984M8	100.2	100.2		
1984M9	93.1	93.1		
1984M10	94.0	94.0		
1984M11	91.9	91.9		
1984M12	128.4	128.4		
1985M1	86.4	85.4	-2.32	-3.44
1985M2	81.4	82.8	-6.94	-5.36
1985M3	97.2	97.4	-4.62	-4.42
1985M4	91.0	90.8	-4.60	-4.81
1985M5	89.4	89.7	-7.26	-7.01
1985M6	93.0	93.8	-5.27	-4.40
1985M7	98.7	97.6	2.86	1.74
1985M8	108.9	105.3	8.62	5.01
1985M9	92.3	95.7	-0.82	2.79
1985M10	95.9	95.9	1.99	1.99
1985M11	97.6	94.3	6.27	2.66
1985M12	121.8	126.4	-5.15	-1.54
1986M1	88.9	89.1	2.88	3.13
1986M2	86.2	86.2	5.88	5.88
1986M3	104.1	102.6	7.13	5.59
1986M4	97.4	98.6	7.08	8.36
1986M5	102.6	99.4	14.79	11.18
1986M6	97.6	101.0	4.98	8.59
1986M7	107.3	107.3	8.74	8.74
1986M8	113.3	114.5	4.02	5.14
1986M9	99.1	98.1	7.36	6.24
1986M10	103.5	103.7	7.96	8.21
1986M11	104.0	104.9	6.56	7.43
1986M12	139.4	138.0	14.47	13.35

(1) Fonte: Banco de Portugal

Tabela 1.5 - Indicador do Consumo Corrente (Hipótese 1)				
Datas	Índice		Taxas de variação homólogas	
	(Base=Jan 1982)			
	Publicado	Corrigido	Publicadas	Corrigidas
1987M1	114.7	95.5	29.09	7.45
1987M2	106.7	91.8	23.81	6.44
1987M3	117.8	106.4	13.17	2.25
1987M4	122.3	104.5	25.57	7.25
1987M5	120.6	105.8	17.50	3.10
1987M6	118.9	103.2	21.82	5.75
1987M7	131.0	120.3	22.01	12.12
1987M8	133.1	122.0	17.55	7.68
1987M9	116.0	107.0	17.10	7.99
1987M10	116.9	104.3	12.99	0.77
1987M11	111.8	107.1	7.47	2.93
1987M12	153.3	142.6	9.98	2.29
1988M1	123.5	116.5	7.63	1.50
1988M2	118.2	109.3	10.78	2.38
1988M3	134.6	121.3	14.30	2.95
1988M4	132.1	125.2	7.96	2.38
1988M5	127.8	125.1	5.99	3.70
1988M6	131.6	125.3	10.64	5.36
1988M7	154.1	138.8	17.68	6.01
1988M8	144.7	136.3	8.69	2.34
1988M9	137.1	127.5	18.15	9.88
1988M10	144.1	124.1	23.22	6.15
1988M11	135.5	115.6	21.24	3.38
1988M12	187.4	158.3	22.25	3.28
1989M1	142.1	128.3	15.10	3.87
1989M2	136.9	122.7	15.80	3.74
1989M3	158.0	144.2	17.38	7.10
1989M4	158.5	141.5	20.00	7.14
1989M5	148.1	132.9	15.86	3.98
1989M6	148.1	135.2	12.59	2.75
1989M7	173.9	158.9	12.85	3.14
1989M8	166.1	153.3	14.79	5.91
1989M9	158.0	141.6	15.26	3.30
1989M10	152.4	146.3	5.76	1.52
1989M11	153.1	143.1	12.95	5.56
1989M12	205.6	194.6	9.68	3.83

Tabela 1.5 - Indicador do Consumo Corrente (Hipótese 1)				
Datas	Índice		Taxas de variação homólogas	
	(Base=Jan 1982)			
	Publicado	Corrigido	Publicadas	Corrigidas
1990M1	158.5	147.6	11.52	3.83
1990M2	153.5	145.0	12.12	5.92
1990M3	171.9	159.4	8.79	0.89
1990M4	164.8	162.0	3.97	2.25
1990M5	153.3	148.2	3.55	0.09
1990M6	163.4	153.3	10.31	3.51
1990M7	191.5	192.7	10.10	10.78
1990M8	169.9	165.9	2.27	-0.16
1990M9	160.3	157.8	1.43	-0.15
1990M10	155.8	150.7	2.26	-1.10
1990M11	157.5	154.8	2.90	1.11
1990M12	214.0	209.7	4.10	2.00
1991M1	192.6	189.9	21.50	19.82
1991M2	157.8	155.5	2.80	1.29
1991M3	196.5	191.7	14.30	11.50
1991M4	166.7	164.2	1.17	-0.37
1991M5	174.3	162.2	13.70	5.75
1991M6	169.8	168.7	3.90	3.22
1991M7	201.2	196.4	5.10	2.56
1991M8	180.4	172.1	6.20	1.29
1991M9	164.8	168.7	2.80	5.22
1991M10	166.4	164.9	6.80	5.78
1991M11	168.9	161.7	7.20	2.67
1991M12	213.8	219.7	-0.10	2.68
1992M1	191.8	190.9	-0.40	-0.89
1992M2	174.1	164.9	10.30	4.45
1992M3	176.8	187.3	-10.00	-4.65
1992M4	182.9	177.7	9.70	6.63
1992M5	187.4	180.3	7.50	3.40
1992M6	171.1	172.8	0.80	1.75
1992M7	211.3	208.8	5.00	3.78
1992M8	185.1	184.1	2.60	2.06
1992M9	166.1	162.0	0.80	-1.68
1992M10	178.6	175.0	7.30	5.11
1992M11	167.7	177.3	-0.70	5.00
1992M12	220.8	221.2	3.30	3.45
1993M1	193.9	191.7	1.11	-0.06

**Anexo de Figuras e Tabelas**

**Segunda Hipótese de Correção**

Tabela 2.1 - Indicador do Consumo Corrente (Hipótese 2)			
Taxas de Variação Homólogas			
Datas	Publicadas	do "Novo" Índice	Corrigidas
1985M1	-2.32	-1.93	-2.48
1985M2	-6.94	-5.05	-2.48
1985M3	-4.62	-4.21	-4.84
1985M4	-4.60	-4.17	-2.00
1985M5	-7.26	-5.36	-6.20
1985M6	-5.27	-3.56	-3.41
1985M7	2.86	2.92	2.89
1985M8	8.62	9.73	5.73
1985M9	-0.82	0.15	3.85
1985M10	1.99	2.82	2.30
1985M11	6.27	7.17	3.99
1985M12	-5.15	-2.31	2.41
1986M1	2.88	3.55	2.71
1986M2	5.88	6.60	6.60
1986M3	7.13	8.43	5.71
1986M4	7.08	7.51	10.36
1986M5	14.79	15.11	11.11
1986M6	4.98	4.04	7.74
1986M7	8.74	7.94	7.42
1986M8	4.02	4.30	5.15
1986M9	7.36	7.35	8.04
1986M10	7.96	7.67	6.83
1986M11	6.56	7.17	7.32
1986M12	14.47	10.91	10.88
1987M1	29.09	17.13	13.13
1987M2	23.81	11.15	11.15
1987M3	13.17	0.28	9.02
1987M4	25.57	13.81	8.25
1987M5	17.50	5.79	6.64
1987M6	21.82	9.87	10.56
1987M7	22.01	11.13	10.29
1987M8	17.55	8.86	8.68
1987M9	17.10	9.20	9.50
1987M10	12.99	10.75	6.75
1987M11	7.47	-0.25	3.45
1987M12	9.98	2.77	2.25

Tabela 2.1 - Indicador do Consumo Corrente (Hipótese 2)			
Taxas de Variação Homólogas			
Datas	Publicadas	do "Novo" Índice	Corrigidas
1988M1	7.63	3.07	3.92
1988M2	10.78	8.24	5.37
1988M3	14.30	13.83	7.14
1988M4	7.96	5.65	8.49
1988M5	5.99	5.66	10.20
1988M6	10.64	7.31	6.07
1988M7	17.68	7.69	4.54
1988M8	8.69	1.91	6.60
1988M9	18.15	7.97	6.31
1988M10	23.22	7.14	7.81
1988M11	21.24	7.12	8.11
1988M12	22.25	14.93	10.09
1989M1	15.10	10.25	14.79
1989M2	15.80	7.35	10.22
1989M3	17.38	10.36	12.52
1989M4	20.00	13.75	10.90
1989M5	15.86	9.20	9.17
1989M6	12.59	4.03	3.91
1989M7	12.85	8.05	7.87
1989M8	14.79	9.77	9.25
1989M9	15.26	8.22	5.04
1989M10	5.76	0.06	4.78
1989M11	12.95	6.60	5.06
1989M12	9.68	4.14	4.99
1990M1	11.52	4.27	4.24
1990M2	12.12	6.67	6.67
1990M3	8.79	9.22	8.35
1990M4	3.97	5.61	6.18
1990M5	3.55	6.56	6.04
1990M6	10.31	11.06	7.88
1990M7	10.10	2.01	6.73
1990M8	2.27	7.08	6.24
1990M9	1.43	5.99	6.14
1990M10	2.26	3.70	3.67
1990M11	2.90	6.14	6.02
1990M12	4.10	6.79	6.61

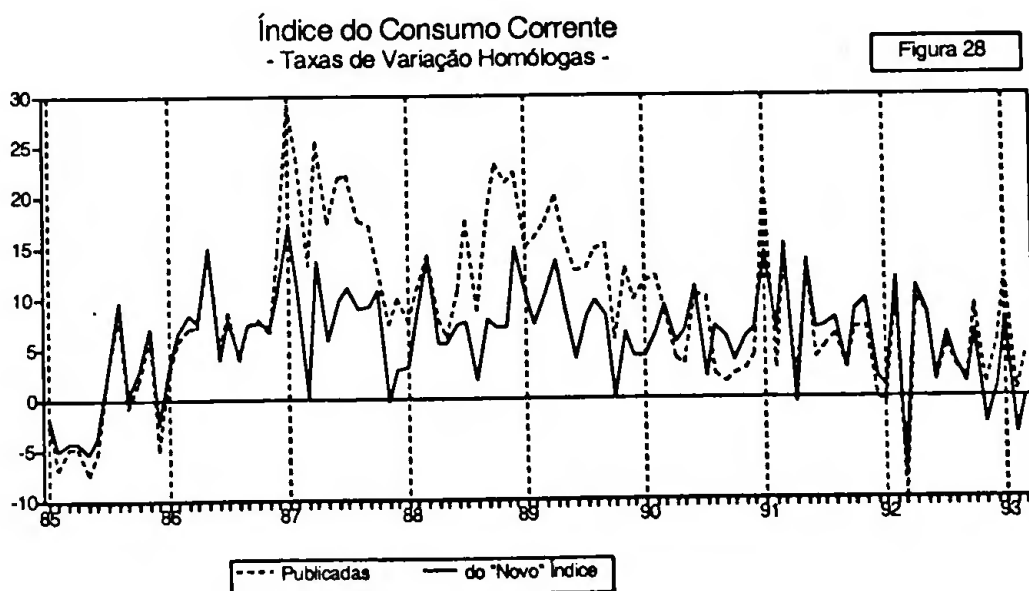
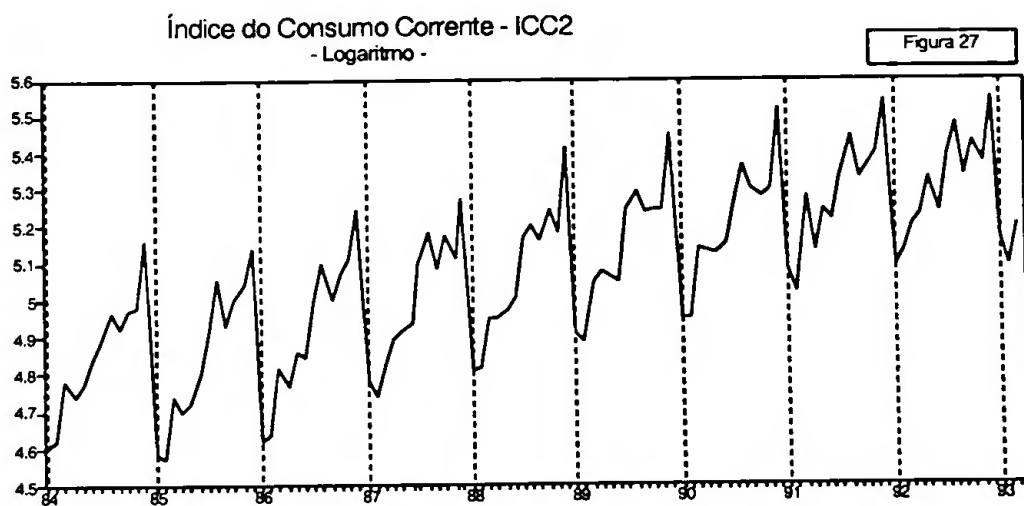
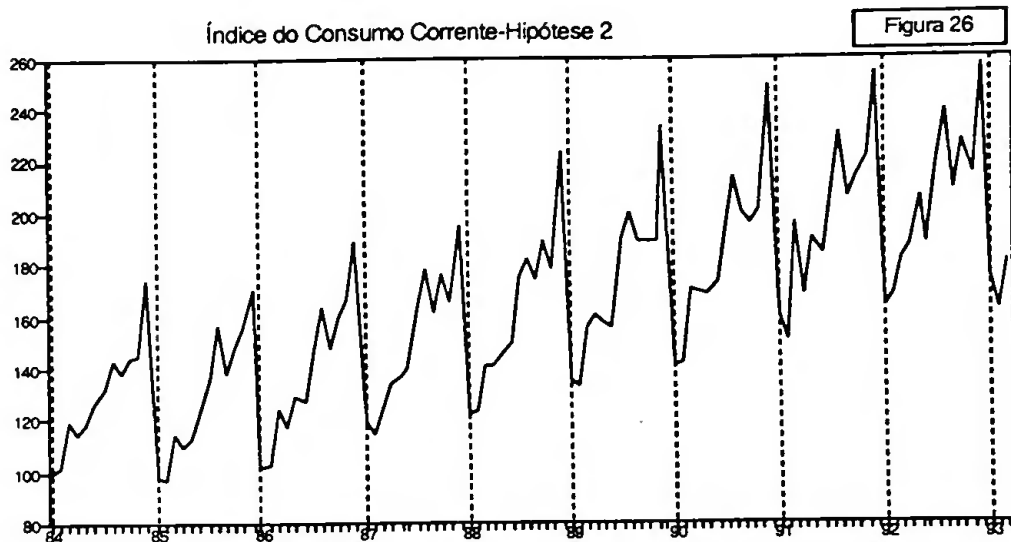
Tabela 2.1 - Indicador do Consumo Corrente (Hipótese 2)			
Taxas de Variação Homólogas			
Datas	Publicadas	do "Novo" Índice	Corrigidas
1991M1	21.50	14.33	13.81
1991M2	2.80	6.45	6.45
1991M3	14.30	15.26	11.58
1991M4	1.17	-0.51	4.71
1991M5	13.70	13.12	12.28
1991M6	3.90	6.75	6.90
1991M7	5.10	7.13	7.10
1991M8	6.20	7.78	3.78
1991M9	2.80	2.85	6.55
1991M10	6.80	8.62	8.10
1991M11	7.20	9.85	6.67
1991M12	-0.10	2.04	6.76
1992M1	-0.40	1.12	0.28
1992M2	10.30	11.91	5.34
1992M3	-10.00	-7.34	1.73
1992M4	9.70	11.00	5.22
1992M5	8.50	8.35	5.20
1992M6	1.50	1.84	6.23
1992M7	5.30	6.28	4.92
1992M8	3.10	3.66	4.33
1992M9	1.80	1.21	2.20
1992M10	9.20	6.45	1.61
1992M11	1.30	-2.74	1.11
1992M12	5.40	1.11	0.56
1993M1	12.00	7.55	4.40
1993M2	0.50	-3.62	2.95
1993M3	4.30	-0.14	-0.17



Tabela 2.2 - Indicador do Consumo Corrente		
	"Novo" Índice (Base=Jan 1984)	"Novo" Índice Corrigido
1984M1	100.0	100.0
1984M2	101.6	101.6
1984M3	119.2	119.2
1984M4	114.1	114.1
1984M5	118.7	118.7
1984M6	126.5	126.5
1984M7	132.2	132.2
1984M8	143.2	143.2
1984M9	137.6	137.6
1984M10	144.0	144.0
1984M11	144.8	144.8
1984M12	174.8	174.8
1985M1	98.1	97.5
1985M2	96.5	99.1
1985M3	114.1	113.4
1985M4	109.3	111.8
1985M5	112.3	111.3
1985M6	122.0	122.2
1985M7	136.1	136.0
1985M8	157.1	151.4
1985M9	137.8	142.9
1985M10	148.1	147.4
1985M11	155.2	150.6
1985M12	170.7	179.0
1986M1	101.5	100.7
1986M2	102.9	102.9
1986M3	123.8	120.7
1986M4	117.5	120.7
1986M5	129.3	124.8
1986M6	126.9	131.4
1986M7	146.9	146.2
1986M8	163.8	165.2
1986M9	147.9	148.9
1986M10	159.5	158.2
1986M11	166.4	166.6
1986M12	189.3	189.3
1987M1	118.9	114.9
1987M2	114.3	114.3
1987M3	124.1	134.9
1987M4	133.8	127.2
1987M5	136.7	137.8
1987M6	139.4	140.3
1987M7	163.2	162.0
1987M8	178.3	178.1
1987M9	161.6	162.0
1987M10	176.6	170.2
1987M11	165.9	172.1
1987M12	194.6	193.6

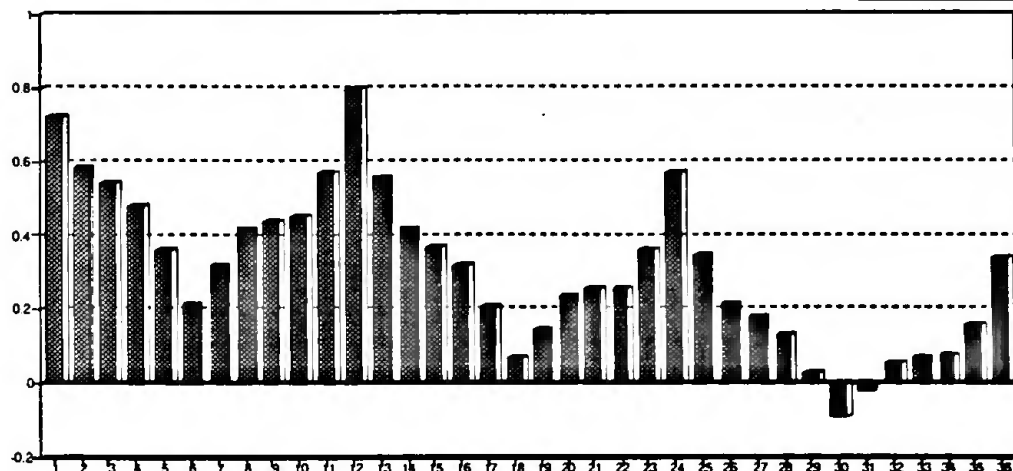
Tabela 2.2 - Indicador do Consumo Corrente		
Datas	"Novo" Índice (Base=Jan 1984)	"Novo" Índice Corrigido
1988M1	122.6	123.6
1988M2	123.8	120.5
1988M3	141.3	133.0
1988M4	141.3	145.1
1988M5	144.5	150.7
1988M6	149.6	147.9
1988M7	175.8	170.6
1988M8	181.8	190.1
1988M9	174.4	171.8
1988M10	189.2	190.4
1988M11	177.8	179.4
1988M12	223.6	214.2
1989M1	135.2	140.7
1989M2	132.9	136.4
1989M3	155.9	159.0
1989M4	160.8	156.7
1989M5	157.8	157.7
1989M6	155.7	155.5
1989M7	189.9	189.6
1989M8	199.5	198.6
1989M9	188.8	183.2
1989M10	189.3	198.3
1989M11	189.5	186.8
1989M12	232.9	234.8
1990M1	140.9	140.9
1990M2	141.7	141.7
1990M3	170.3	168.9
1990M4	169.8	170.7
1990M5	168.1	167.3
1990M6	172.9	167.9
1990M7	193.7	202.7
1990M8	213.6	212.0
1990M9	200.1	200.4
1990M10	196.3	196.3
1990M11	201.1	200.9
1990M12	248.7	248.3

Tabela 2.2 - Indicador do Consumo Corrente		
Datas	"Novo" Índice (Base=Jan 1984)	"Novo" Índice Corrigido
1991M1	161.1	160.4
1991M2	150.9	150.9
1991M3	196.3	190.0
1991M4	168.9	177.8
1991M5	190.2	188.8
1991M6	184.5	184.8
1991M7	207.5	207.5
1991M8	230.3	221.7
1991M9	205.8	213.2
1991M10	213.3	212.2
1991M11	220.9	214.6
1991M12	253.8	265.5
1992M1	162.9	161.6
1992M2	168.8	158.9
1992M3	181.9	199.7
1992M4	187.5	177.7
1992M5	206.1	200.1
1992M6	187.9	196.0
1992M7	220.6	217.7
1992M8	238.7	240.2
1992M9	208.3	210.3
1992M10	227.0	216.7
1992M11	214.9	223.4
1992M12	256.6	255.2
1993M1	175.2	170.1
1993M2	162.7	173.8
1993M3	181.6	181.6



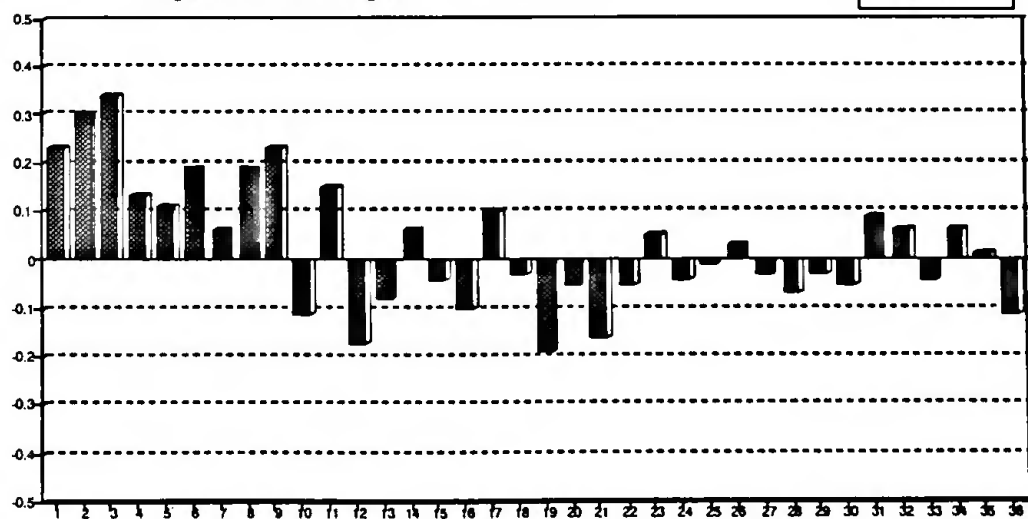
Função de Autocorrelação de LICC2

Figura 29

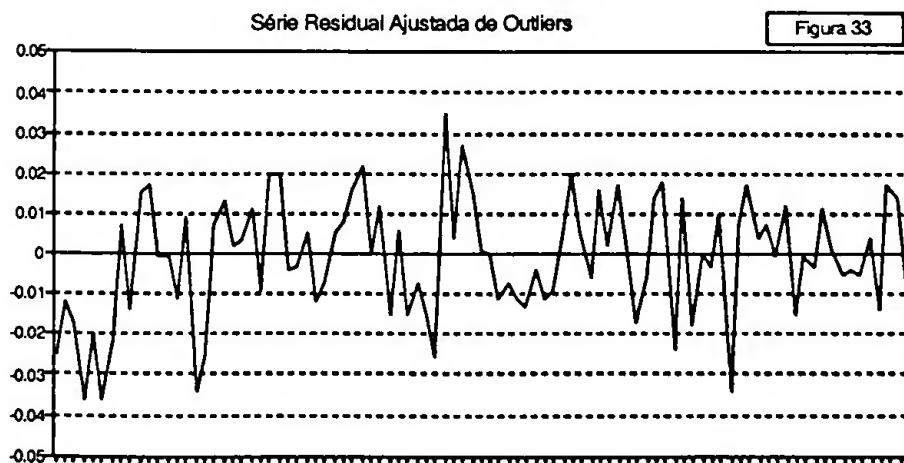
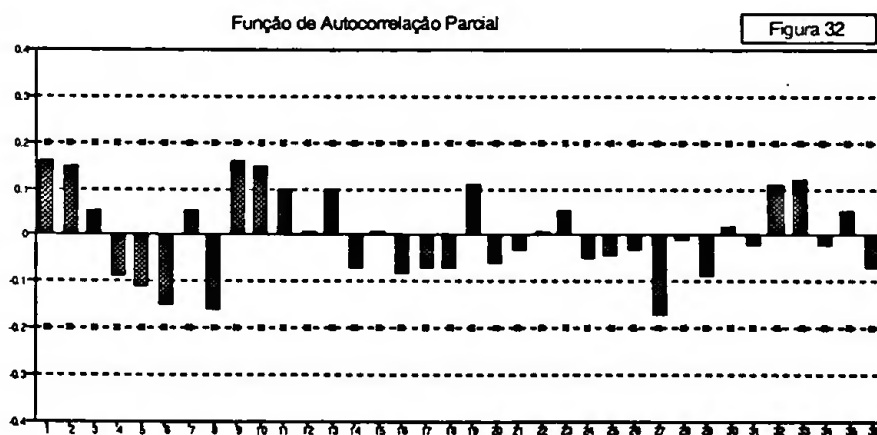
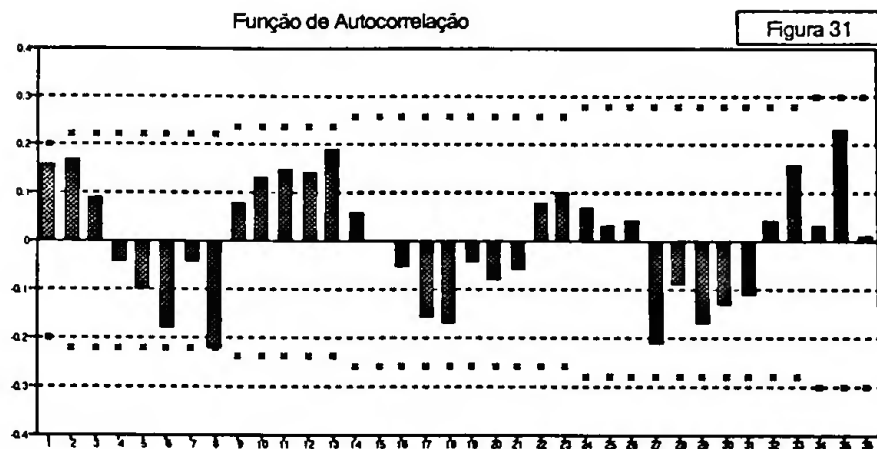


Função de Autocorrelação de  $(1 - B^{12})LICC2$

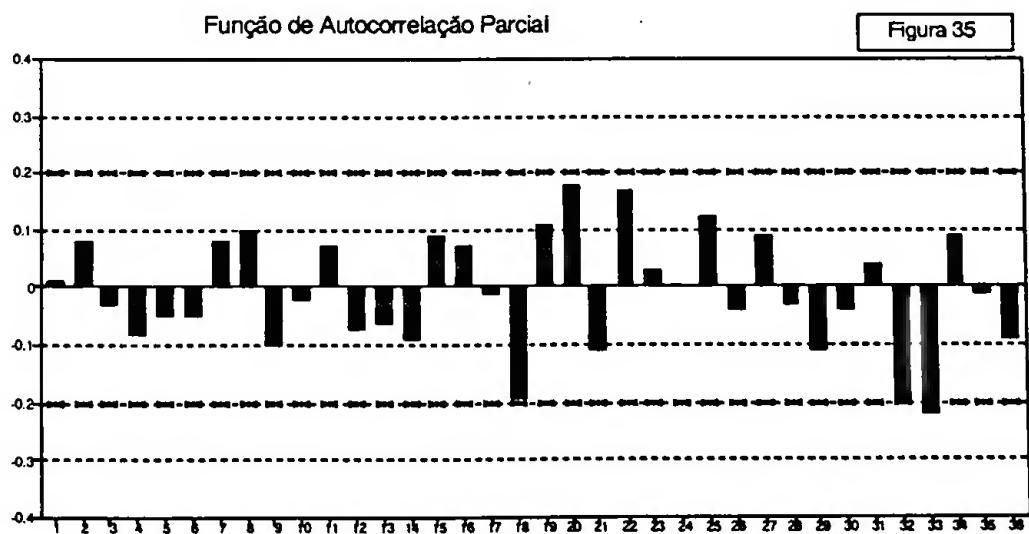
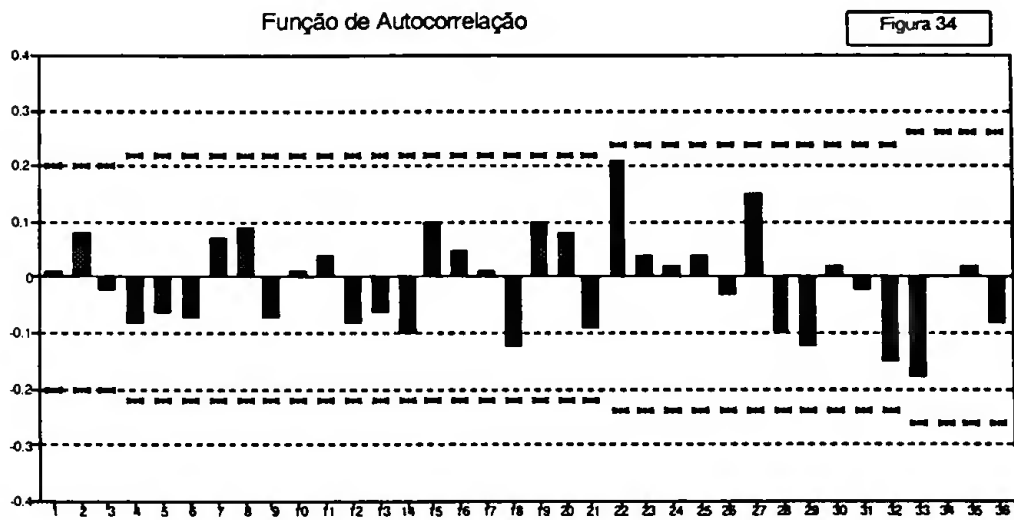
Figura 30



# SÉRIE RESIDUAL DO MODELO (4.12)



## SÉRIE RESIDUAL DO MODELO "FINAL"



Série Residual Ajustada de Outliers

Figura 36

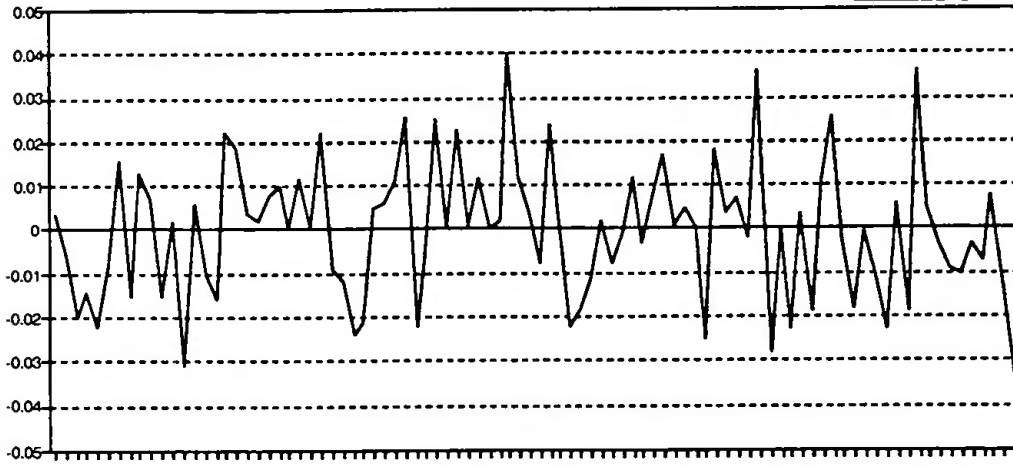


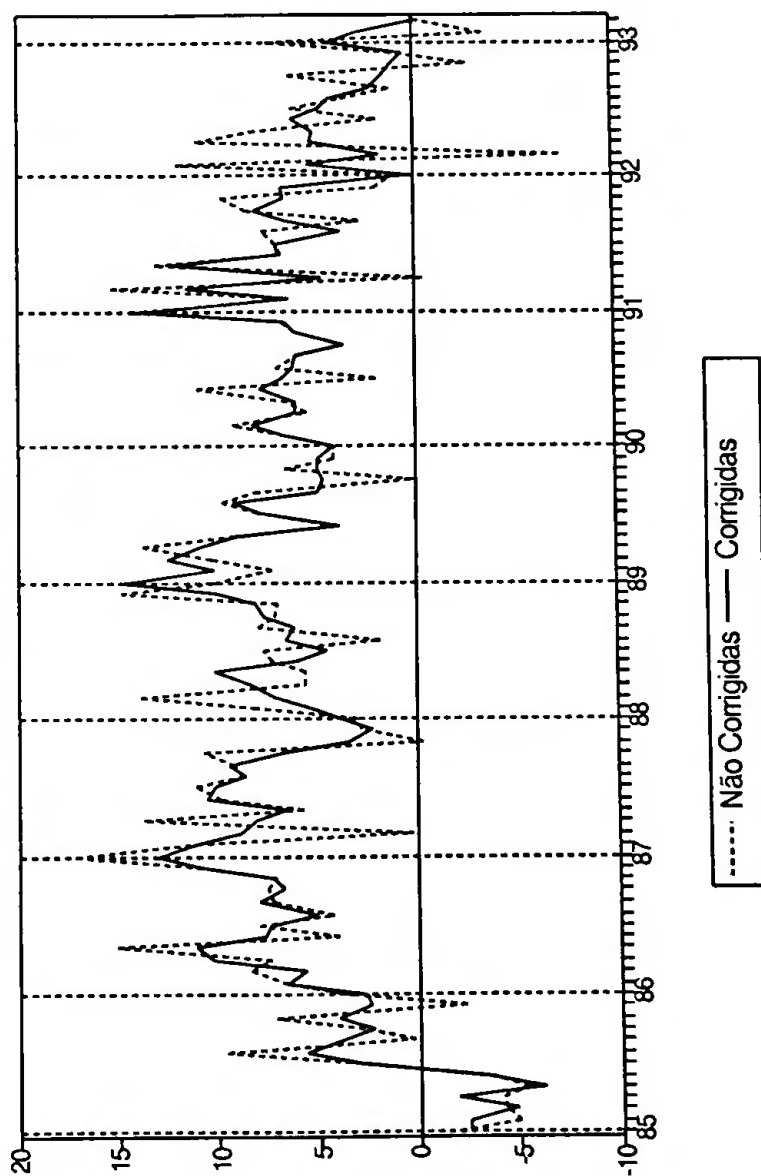
Tabela 2.3.D - Coeficientes de Correlação das Estimativas dos Parâmetros

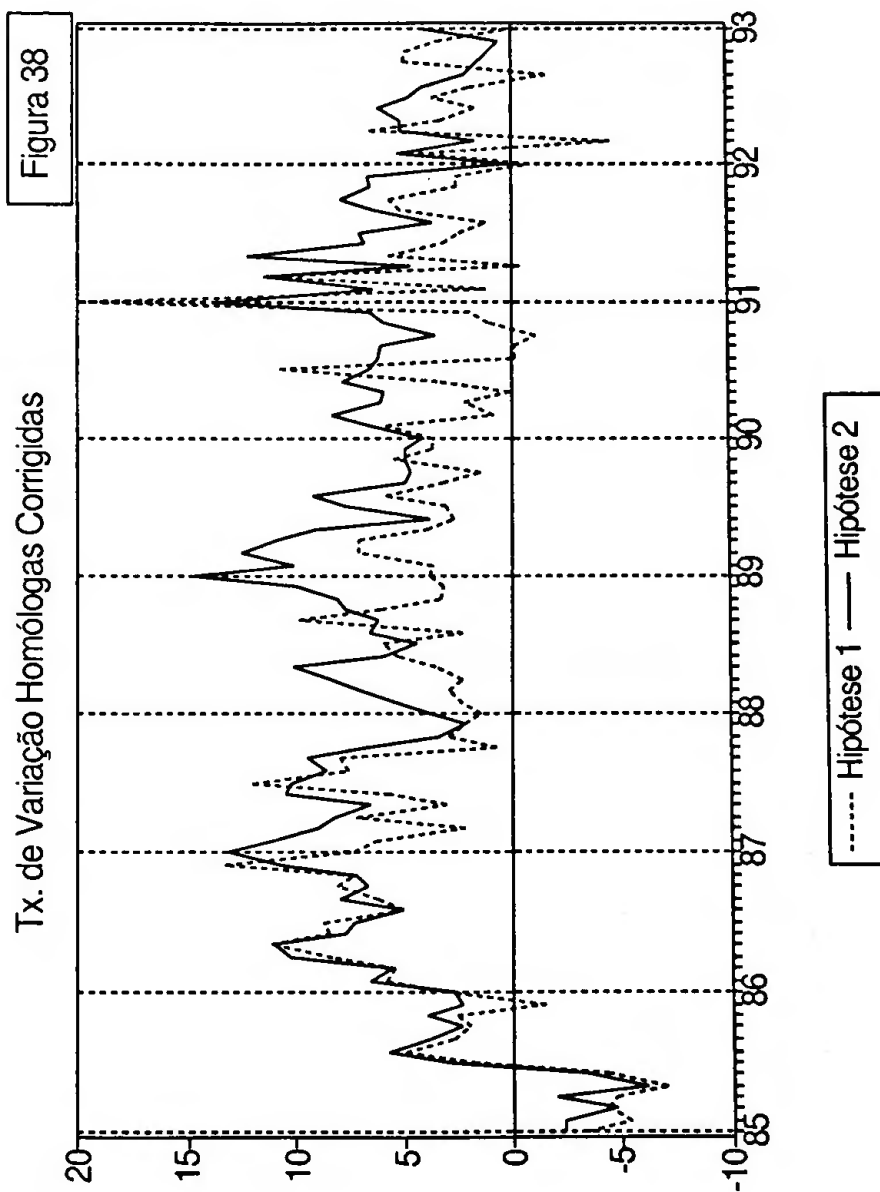
	1	2	3	4	5	6	7	8	9	10	11	12	13
1	1												
2	0	1											
3	0	.73	1										
4	0	.86	.81	1									
5	0	.87	.83	.82	1								
6	0	.79	.87	.86	.77	1							
7	0	.85	.82	.83	.88	.71	1						
8	0	.76	.88	.87	.85	.91	.76	1					
9	-.01	.01	-.13	-.19	-.14	.04	-.19	-.09	1				
10	0	-.03	.25	.17	.13	.02	.20	.10	-.87	1			
11	0	.08	.11	.15	.10	.09	.06	.10	-.40	.34	1		
12	.01	-.10	-.06	-.10	-.08	-.09	-.05	-.06	-.07	.09	.04	1	
13	.04	.10	.09	.10	.08	.11	.08	.09	-.04	.02	.03	-.02	1

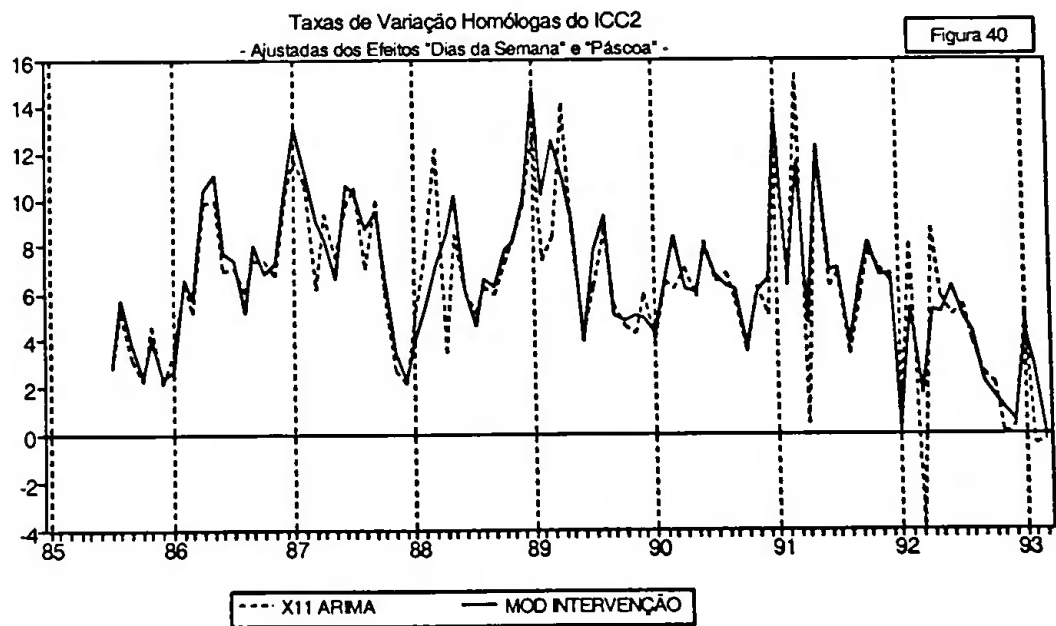
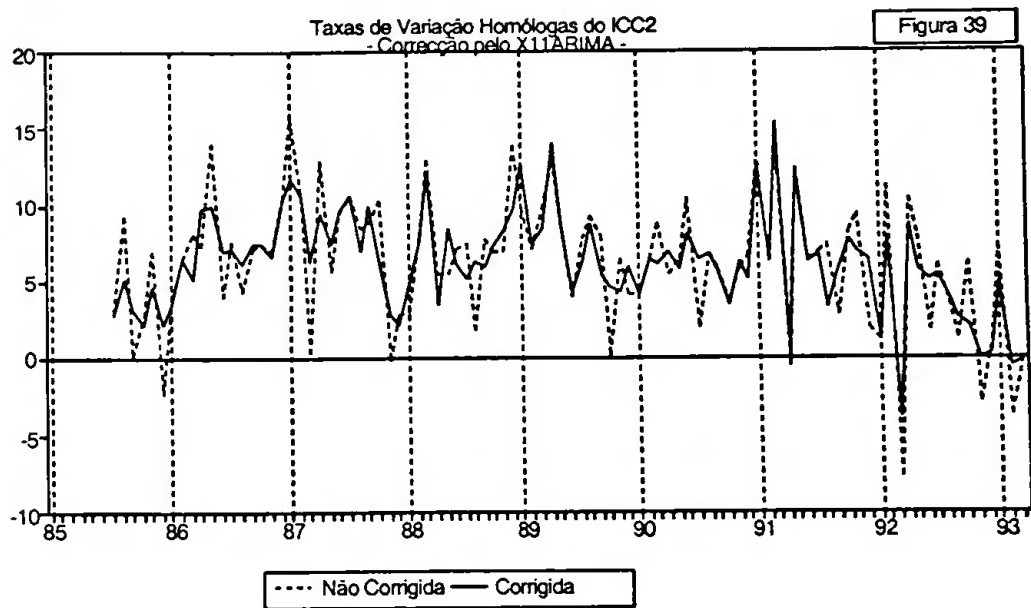


Índice do Consumo Corrente - Hipótese 2  
- Taxas de Variação Homólogas -

Figura 37







## Referências Bibliográficas

- Box, G. e Tiao, G. (1975), "Intervention Analysis with Applications to Economic and Environmental Problems", *Journal of American Statistical Association*, vol.70, number 347.
- Chang, I., Tiao, G.C. e Chen, C. (1988), "Estimation of time series parameters in the presence of outliers", *Thecnometric*, 30, nº 2, 193-204.
- Chen, C., Liu, L. e Hudak, G., "Outlier Detection and Adjustment in Time Series Modeling and Forecasting", working paper, SCA.
- Chen, C. e Liu, L. (1990), "Recent Developments of Time Series Analysis in Environmental Impact Studies", paper.
- Chen, C. e Liu, L. (1993), "Joint estimation of model parameters and outlier effects in time series", *JASA*, vol.88, nº 421, pp.284-297.
- Chen, C. e Liu, L. (1993), "Forecasting Time Series with outliers", *Journal of Forecasting*, 12, 13-35.
- Cleveland, W.S. e Devlin, S. (1980), "Calendar effects in monthly time series: detection by spectrum analysis and graphical methods", *JASA*, Sept., vol.75, nº 371.
- Cleveland, W.P. (1986), "Calendar adjustment and time series", nº 198, *Division of Research and Statistics, FED, Washington*.
- Dagum, E.B. (1988), "The X-11 ARIMA/88 Seasonal Adjustment Method - Foundations and User Manual", *Time Series Research and Analysis Division, Statistique Canada*.

- Dias, F. (1987), "Um indicador do consumo corrente para Portugal - revisão", *Boletim Trimestral*, Março, Banco de Portugal.
- Espasa, A. (1982), "El comportamiento de series economicas: movimientos atipicos y relaciones a corto y largo plazo", *Banco de Espanha*, Working paper 8201.
- Espasa, A. (1990), "Univariate methodology for short term economics", *Banco de Espanha*, Working paper 9003.
- Escrivá, J.L. e Sastre, M.T. (1988), "La desestacionalización de los agregados monetarios", *Boletín Económico*, Março, *Banco de Espanha*.
- Fox, A.J. (1972), "Outliers in time series", *J. of Royal Statistics Soc., Ser. B*, 43, 350-363.
- Glass, G.V. (1972), "Estimating the effects of intervention into non-stationary time series", *American Educational Research Journal*, nº 3, 463-477.
- Harvey, A. (1990), "Data irregularities", *Forecasting, Structural Time Series Models and the Kalman Filter*, cap. 6, section 4, Cambridge Press University.
- Hillmer, S. C. (1984), "Monitoring and adjusting forecasts in presence of outliers", *Journal of Forecasting*, 3, 205-215.
- Hudak, G. e Liu, L., "Transfer Function Modeling: Simplification and Applications", working paper, SCA.
- Krishnamurthi, L., Narayan, J. e Raj, S.P. (1989), "Intervention analysis using control series and exogenous variables in a transfer function model: A case study", *International Journal of Forecasting*, 5, pp. 21-27.
- Ledolter, J. (1988), "The effects of additive outliers on the forecasts from ARIMA models", *International Journal of Forecasting*, 5, 231-240.

- Liu, L. (1986), "Identification of time series models in the presence of calendar variation", *International Journal of Forecasting*, 2, 357-372.
- Liu, L. (1991), "Dynamic relationship analysis of US gasoline and crude oil prices", *Journal of Forecasting*, 10, 521-547.
- Liu, L. e Liu (1991), "Forecasting residential consumption of natural gas using monthly and quarterly time series", *International Journal of Forecasting*, 7, 3-16.
- Liu, L., Hudak, G., Box, G, Muller, M e Tiao, G (1992), "Intervention analysis", *The Sca Statistical System: Reference Manual for Forecasting and Time Series Analysis Using the SCA System*, Scientific Computing Associates, vol.1, cap.6
- Liu, L., Hudak, G., Box, G, Muller, M e Tiao, G (1992), "Outlier Detection and Adjustment", *The Sca Statistical System: Reference Manual for Forecasting and Time Series Analysis Using the SCA System*, Scientific Computing Associates, vol.1, cap.7
- Mills, T. (1990), "Intervention analysis and the detection of outliers", *Time Series Techniques for Economists*, chapter 12, pp.235-247, Cambridge Press University.
- Mills, T.(1991), "Estimating trend growth rates of the U.K. Monetary Aggregates", *Journal of Forecasting*, vol.10, 269-283.
- Morales, E., Espasa, A., Rojo, M.L. (1990), "Univariate methods for the analysis of the industrial sector in Spain", Documento de Trabalho nº 8904, *Banco de Espanha*.
- Murteira, B., Muller, D. e Turkman, K. (1993), "Modelos de Função Transferência e de Intervenção", *Análise de Sucessões Cronológicas*, cap. 9, McGraw-Hill.
- Muller, D. (1990), *Análise de Sucessões Cronológicas: Modelos de Função Transferência. Teoria e Prática*, CEMAPRE, ISEG.



- Santos, E. (1986), "Um indicador do consumo corrente para Portugal", *Boletim Trimestral*, Junho, Banco de Portugal.
- Thury, G. (1988), "Intervention analysis of consumer expenditure in Austria", *Empirica*, 15(2), pp.295-325.
- Tiao, G.C. (1985), "Autoregressive moving average models, intervention problems and outlier detection in time series", *Handbook of Statistics*, vol.5, pp. 85-118.
- Tsay, R.S. (1986), "Time series model specification in the presence of outliers", *JASA*, 81, 132-141.
- Tsay, R.S. (1988), "Outliers, level shifts and variance changes in time series", *Journal of Forecasting*, 7, 1-20.
- Wei, W. (1990), "Intervention Analysis and Outlier Detection", *Time Series Analysis - Univariate and Multivariate Methods*, cap. 9, Addison-Wesley.